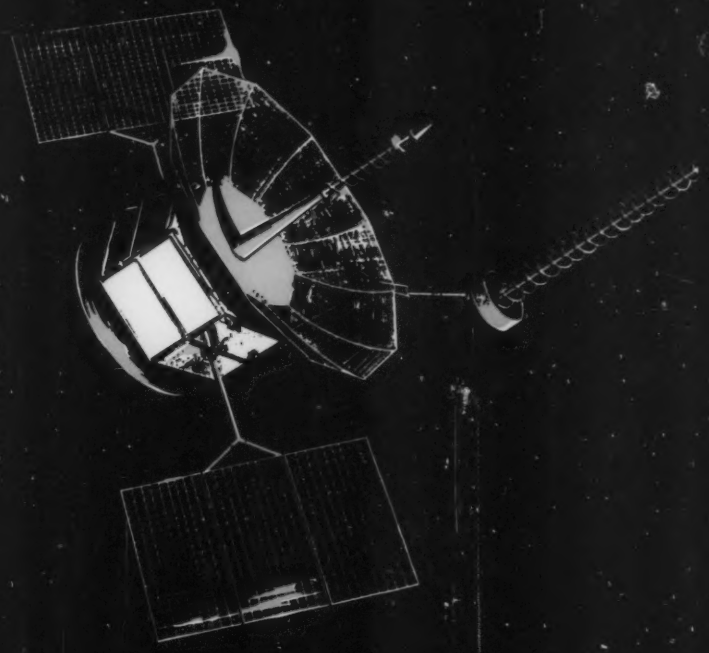
Naval Research Reviews

Office of Naval Research
Fall/Winter 1980/1981
Vol XXXIII No. 1



The giant radio located in the Owens Valley, 260 mile from Los Angeles, was built in 1959 by the California Institute of Technology with funds from the Office of Naval Research. During the first two months of operation the telescope located and identified nine new radio sources outside our galaxy. Working in tandem, the twin dishes produced a resolving power greater than any radio telescope of the day.

During the formative years of radio astronomy, ONR was the major sponsor in the United States. The Owens Valley telescope was one of several built and operated with ONR funds during the 1950's through the 1960's.

Naval Research Reviews

Office of Naval Research
Fall/Winter 1980/1981
Vol XXXIII No.1

Chief of Naval Research	RADM A. J. Baciocco, Jr., USN
Technical Director	Dr. J. A. Smith
Scientific Editor	Dr. E. I. Salkovitz
Associate Scientific Editor	Dr. G. A. Neece
Associate Scientific Editor	Dr. J. T. Lester
Associate Scientific Editor	Dr. D. A. Patterson
Managing Editor	M. J. Whetstone
Chief Writer/Editor	W. J. Lescure, III
Art Editor	Edward Bailey

ARTICLES

2 The Environment & Earth-Space Propagation
by John M. Goodman

20 Study of Iron in Magnetotactic Bacteria
by Richard B. Frankel and Richard P. Blakemore

24 The Acoustic Microscope:
by C. F. Quate

34 Frequency Independent Beamforming
by Robert L. Sternberg and Warren A. Anderson

48 Components of Causal Inference
by Robert J. Sternberg and Miriam W. Schustack

68 A New Propulsor Concept for Ships
by Herman E. Sheets

DEPARTMENTS

33 Profiles in Science
Dr. Alan J. Heeger

63 Research Notes

About Our Cover

Fleet Satellite Communications Systems (FLTSATCOM) will provide high-priority UHF communications between naval aircraft, ships, submarines, groundstations, and Presidential Command Networks. Four satellites, such as the one depicted on the front cover with the erupting sun, will provide complete earth coverage except for polar regions. On the back cover is Defense Satellite Communication System Phase II (DSCSII) orbiting earth. DSCSII provides the highest transmitted power to date. See article beginning on page 2. Photograph courtesy of Rockwell International.

Naval Research Reviews publishes articles about research conducted by the laboratories and contractors of the Office of Naval Research and describes important naval experimental activities. Manuscripts submitted for publications, correspondence concerning prospective articles, and changes of address, should be directed to Code 733, Office of Naval Research, Arlington, Va. 22217. Requests for subscription should be directed to the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Yearly subscription prices for this quarterly publication are $6.50 in the United States and Canada and $8.15 elsewhere. The issuance of this periodical is approved in accordance with Department of the Navy publications and printing regulations.

The Environment and Earth-Space Propagation –Challenges for the Future

by

John M. Goodman

Naval Research Laboratory

Introduction

The advantages of utilizing space for telecommunications are well known in both the commercial and military arenas. A small complement of satellites at synchronous orbit, for example, may provide nearly global coverage and may be designed to support small disadvantaged customers as well as those possessing large antenna structures and sophisticated acquisition and processing capabilities. Modern navigational and timing needs can also be satisfied through exploitation of space platforms, and NAVSTAR/GPS is a system that exemplifies the utilization of space for those purposes. Applications of space in surveillance and kindred areas also exist, and spaceborne instruments for monitoring the exoatmospheric environment and transmissions from the sun abound. There is an obvious charm in the utilization of space for various purposes, but it must be tempered with the realization that most applications require the transmission of intelligence or data between space platforms and other space segments or a ground terminal. Thus the channel or the propagation path clearly becomes an integral part of the total system. Its nuisance value as a perturbation source depends on the extent to which it differs from free space at a specified frequency.

Since the advent of the space age, there has been an increasing awareness of the benefits that might accrue from utilization of orbiting systems for a variety of purposes. Because of the unique vantage point provided by space, the potential for military and commercial communication, navigation, surveillance, earth observation, and space research has been increased significantly. The technologies for the successful launch, orbital maintenance, and operation of highly sophisticated systems have been developed over the years since 1958; and currently there are more than 800 payloads in orbit of which approximately 30% are still actively performing their assigned missions. Of these, as reported in a popular military journal, approximately 75 are active communication satellites with the majority being Soviet systems.[1] These statistics, whether precise or not, do indicate the increased emphasis being placed upon space systems by the major industrial nations to fulfill national objectives in the C^3I arena.

In order to utilize the benefits of space fully, it is necessary to account for the subtle environmental factors that may continuously interfere with successful operations or may constrain or limit the performance of the system at seemingly random epochs in time. In principle, an *a priori* knowledge of the full range of problems imposed by nature should lead to the design of robust systems (including both space segments and earth terminals) that either are impervious to disturbances or may adapt to the changes in some sense. At the very least, the environmental knowledge is a prerequisite to intelligent system design. Otherwise "band-aid" approaches will be dictated after-the-fact. For the disadvantaged user who is constrained by operational environment, cost, real-estate, or other factors, the entire

This antenna has a 60 dB gain and is located at major SATCOM stations.

Dr. John M. Goodman is Head of the Ionospheric Effects Branch of the Space Science Division at the Naval Research Laboratory. He has many publications in the area of ionospheric physic and radiowave propagation in the ionospheric environment and has edited two volumes associated with International Ionospheric Effect Symposia for which he served as chairman in 1975 and 1978.

burden of robustness must be borne by the space segment and this cost may be too large. Techniques of diversity (including frequency, polarization, time or coding, and space) and various resource management schemes involving re-routing and/or gateway scenarios are options that have been explored for mitigation or avoidance of environmental effects. In some cases, the implementation of these techniques is either too costly or cumbersome considering the perceived risk.

Thus, the growing utilization of space has increased the importance of understanding the intervening environment, including the troposphere and its weather, the ionosphere, the magnetosphere, and the interplanetary medium. Serious basic research efforts have been initiated over the years to ascertain the basic nature of the total environment or at least its statistical climatology, with the ultimate goal being to estimate the range and variance of deleterious system effects. This information is utilized in system design and development, but may also be of value in developing forecasting and assessment techniques for use in operational resource management. The next section will outline some of the constraints placed upon space systems that rely upon the earth-space propagation path.

A Resume of Earth-Space Radio Propagation Effects

One of the earlier papers dealing with this topic was due to Millman[2] who surveyed both tropospheric and ionospheric effects on radio propagation between the earth and space vehicles. A basic paper with special emphasis on the ionospheric aspect of earth-space propagation by Lawrence et al[3] is also noteworthy although it does not include the subsequent important developments in scintillation morphology. For background the reader is referred to books such as those by Davies,[4] Kelso,[5] and Budden,[6] and to the excellent library of the Advisory Group for Aerospace Research and Development (AGARD) publications that have treated various propagation effects over the years. Good introductions to propagation effects due to the lower atmosphere are presented in texts by Kerr,[7] Battan,[8], and Bean and Dutton,[9] and the author is fortunate to have unpublished lecture material on millimeter-wave propagation effects provided by Ippolito.[10] Numerous papers have been published over the years dealing with the material covered in this article. Especially noteworthy are recent papers by Rush[11] on ionospheric radio propagation, Klobuchar[12,13] on ionospheric time delay, Aarons[14] on scintillation morphology, Fremouw and Rino[15] on scintillation modelling and statistics, Crane[16] on prediction of rainfall effects and Bean and Dutton[17] as well as Boithias[18] on tropospheric propagation. A sketch of the ionospheric effects upon earth-space propagation is available in a CCIR report.[19] Additionally the proceedings of the two previous NRL-sponsored Ionospheric Effects Symposia (IES),[20,21] and a recent International Committee on Space Research symposium[22] are compilations of recent research efforts and provide the user of environmental data with a view of the opportunities for improvement in system design and operation.

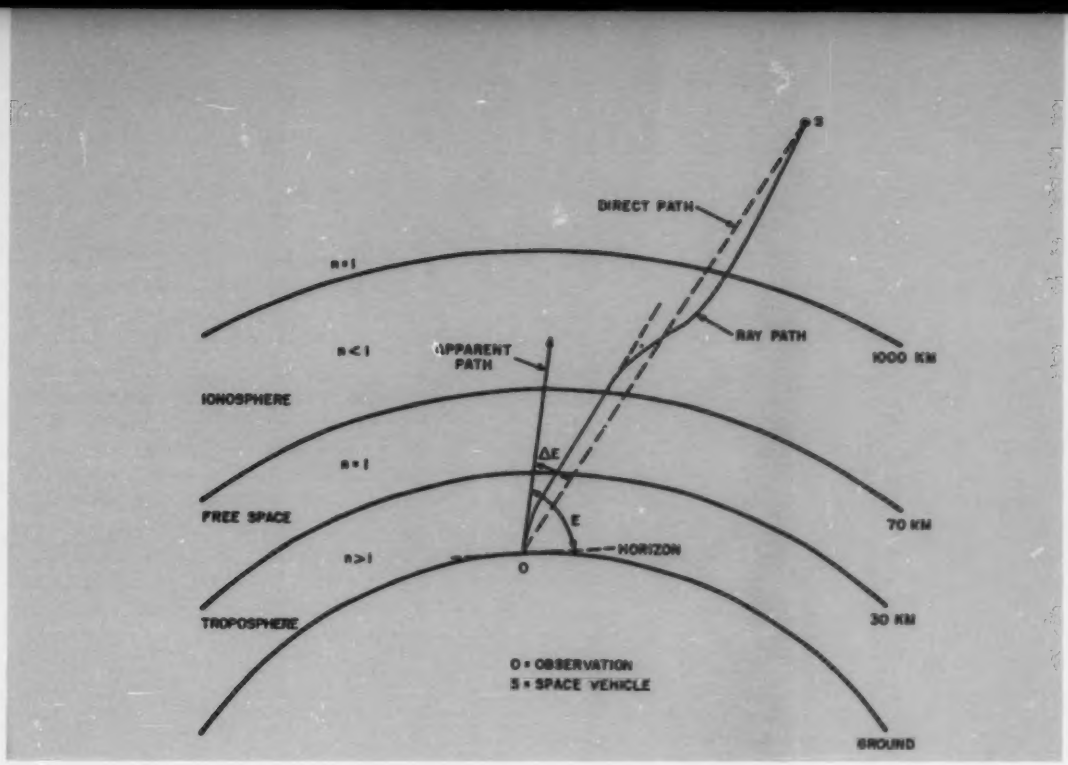


Figure 1. Radiowave Trajectory through the Troposphere and Ionosphere.

There is a myriad of radio propagation effects that come to mind as one considers the nature of media through which rays must propagate between earth and space. The obvious effects of refraction and absorption in the lower atmosphere have their counterparts in the ionized upper atmosphere, but there are additional effects in the ionosphere due to the magnetoionic medium. The Faraday rotation of linearly polarized radiowaves and various differential effects resulting from bi-refringence come to mind. Although tropospheric weather (and the propagation effects it generates) is not uncomplicated, the primary concern vis-a-vis earth-space propagation has traditionally been the ionosphere and its personality. This was so because space frequencies were lower near the advent of the space age than they are now or will likely be in the future. The following sections discuss some of the more important RF propagation effects for rays that traverse both the troposphere and the ionosphere.

Refraction in Earth-Space Propagation

Figure 1 depicts both the tropospheric and ionospheric components of bending introduced over an earth-space path. This is due, of course, to the non-vanishing refractivity in the earth-space medium. The earth-space refractivity may be written

$$N(s) = (n-1) \times 10^6 = N_t + N_I \qquad (1)$$

$$= \frac{77.6}{T(s)} \left[p(s) + \frac{4810\, \epsilon(s)}{T(s)} \right]$$

$$- \frac{40.28 \times 10^{-6} N_e(s)}{f^2}$$

where N(s) is the refractivity, n is the refractive index, N_T and N_I are the tropospheric and ionospheric components of refractivity, T(s) is the air temperature (^oK), p(s) is the atmospheric pressure (mb), ϵ(s) is the partial vapor pressure (mb), f is the radiofrequency (MHz), N_e is the electron density (M^{-3}), and s is the distance parameter.

Figure 2 depicts a typical refractivity profile from ground level through the ionosphere. It is noteworthy that $N_t(h)$ [where h replaces s for zenithal propagation] is independent of f whereas $N_I(h)$ is inversely proportional to the square of f. Thus, for the higher space frequencies, the tropospheric refractivity dominates the ionospheric component. This is not to say, however, that ionospheric refractivity effects can be ignored at higher space frequencies.

With respect to gross tropospheric bending, Bean and McGavin,[23] Bean and Cahoon,[24] and Bean et al[25] have shown that surface values of the refractivity N may

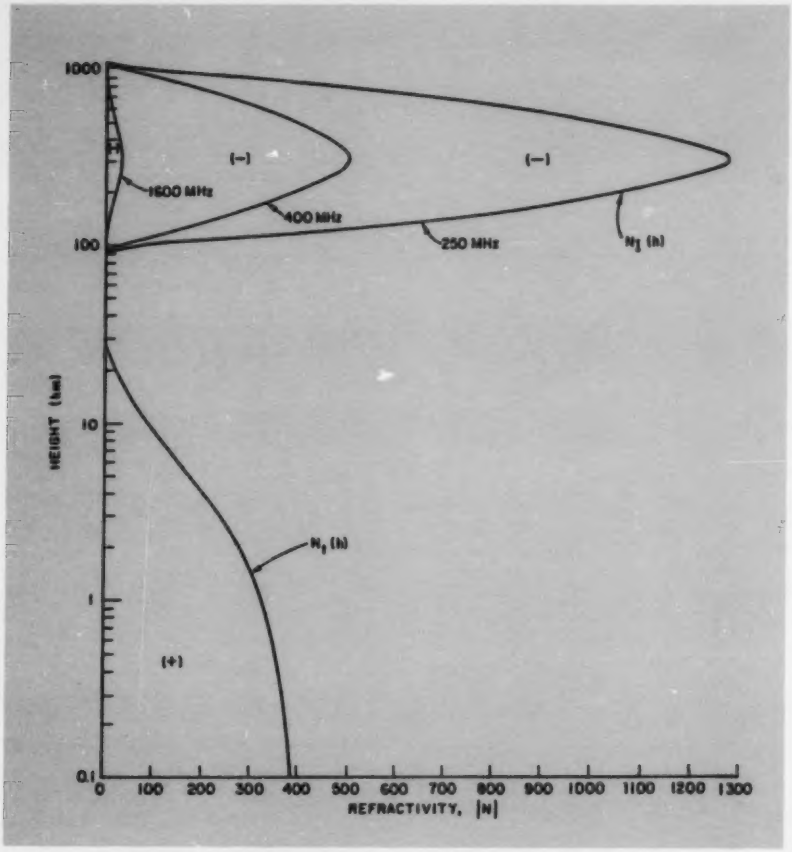


Figure 2. Radio refractivity versus altitude. Representative ionosphere profiles are shown for 250, 400, and 1600 MHz.

be used to approximate the effect. A considerable amount of work has been done by Bean and his colleagues as well as others in recent years. For example the reader should examine material in the NATO/AGARD conference on Tropospheric Radiowave Propagation[26] which contains a review by Millman[27] (also see references 17 and 18). Bending through the troposphere may be given by the following approximate relation:[23]

$$\tau_{\infty}(\theta) = b(\theta)N_s + a(\theta) \qquad (2)$$

where θ is the initial ray elevation angle, N_s is the surface refractivity, and a and b are constants.

It is noteworthy that a climatology for the surface refractivity does exist and models for $N_s(h)$ also have been published.[17] The simple recipe given by equation 2 is obviously insufficient under extreme climatic conditions, and in particular the "humid" term (i.e., 4810 e/T in equation 1) exhibits large variations. It is well known that both the horizontal and vertical details of refractivity are of great importance in radio propagation. If the vertical gradient of N has a very large negative value (for example ≤ -157 N units per km), then ducting will occur. If the vertical gradient is positive (or slightly negative), then super refraction will occur. Such matters, however, are generally of little significance to earth-space links except at very low propagation angles. It is clear from this discussion and also from Figure 2 that the refractivity of the atmosphere is not a pure exponential. There have been attempts to construct both 2-part and 3-part exponential models,[28] but even these sophisticated models are inadequate to explain the situation in the tradewind zones.

The index of refraction in the troposphere is greater than unity and since dN/dh is negative in that region, non-zenithal rays are bent away from the normal. As a result, the apparent elevation to the space object is

higher than its actual value and the radio range to the object is larger than the geometrical distance. The index of refraction in the ionosphere is less than unity. Nevertheless, the net effect of the ionospheric component of refraction is to combine with the tropospheric contribution such that the total refraction is essentially the sum of the two components. It is noteworthy that one component of range error is introduced by the combined effect of tropospheric and ionospheric refraction. However, the total range error results from both a time delay due to an increase in path length (caused by bending) and a reduction in radiowave signal velocity. The time delay component of range error dominates for elevation angles above 3°.

In summary, the ionospheric component of refractivity remains quite important in comparison to the tropospheric component in the UHF band for elevation angles above a critical value, say 5°. Below this elevation, the tropospheric component begins to dominate and other effects such as ducting, atmospheric multipath, and atmospheric scintillation become increasingly important. Millman[2] reviews many of these effects. The tropospheric errors are typically less variable at higher elevation angles, and the climatology of surface refractivity and its variance exhibits a degree of stability (or may be readily measured). As a result, routine subtraction of tropospheric bending and range bias may be accomplished with some degree of confidence. Ionospheric errors are not as easy to address since the climatology is not well known and the variances are relatively large and moderately unpredictable. This situation persists despite valiant attempts at ionospheric prediction and assessment that continue to this day.

Attenuation in Earth-Space Paths

The absorption introduced by the ionosphere is of negligible significance at UHF and higher space frequencies. The largest effects would occur during polar cap absorption (PCA) and kindred auroral events. Even these extreme events will produce absorption of less than 1 dB at space frequencies above VHF. The interested reader is referred to AGARDOGRAPH No. 53,[29] Davies,[4] and CCIR Report 263-3.[19]

The troposphere, on the other hand, introduces potentially severe attenuation of earth-space signals. Attenuation, including both absorption and scattering, depends on the various atmospheric constituents and their frequency-dependent characteristics. Figure 3 depicts the specific attenuation for normal absorption, resonance absorption, rain, and fog for a horizontal path. Absorption and scattering typically are unimportant for frequencies less than 3 GHz. The resonance absorption effects are particularly important above 3 GHz. For an elevation angle of 45°, the attenuation exceeds 100 dB over a typical earth-space path at 61, 119, 183, and 324 GHz. Scattering from clouds and hydrometeors (rain) will reduce the amount of energy available in the forward direction, will change the polarization of the radiowaves, and will deflect energy back toward the transmitter as well as in other directions.

Since the resonance absorption peaks of H_2O and O_2 are well known and can be avoided, the most significant future problem for earth-space propagation at frequencies between 10 and 100 GHz is likely to be rainfall attenuation. A summary of propagation impairments on 10-100 GHz satellite links reviews the topic with emphasis upon techniques for system design.[30] Two types of rain are considered: stratiform and convective. Stratiform rain generally develops in stable masses of air, along frontal surfaces typically, and is characterized by steady and uniform rain over wide areas for hours to days; low rain rates are typical. Convective rainstorms, on the other hand, develop in highly unstable air masses, are limited in extent and duration and are characterized by high rain rates. The specific attenuation α by hydrometeors is given by

$$\alpha\,(\mathrm{dB/km}) = 4.343 \int_0^\infty Q_T(r)\,\eta\,(r,R)\,dr \qquad (3)$$

where Q_T is the attenuation cross section/drop, η is the drop size distribution, r is the drop radius, and R is the rain rate.

The attenuation function Q_T is determined from classical Mie and/or Rayleigh scattering theory and for computation purposes, several drop size distributions are available.[31,32,33] Nevertheless, there is a convenient empirical relationship between rain rate R and α:[34,35]

$$\alpha\,(\mathrm{dB/km}) = aR^{b} \qquad (4)$$

where a, b are frequency-dependent constants.

Thus, given an estimate of the rain rate R, the specific attenuation is readily determined by equation 4 since curves of a (f) and b (f) are prescribed.[36] Uniform rain attenuation is proportional to the cosecant of the ray elevation angle since the total attenuation A is proportional to the product of the specific attenuation α and the path length L.

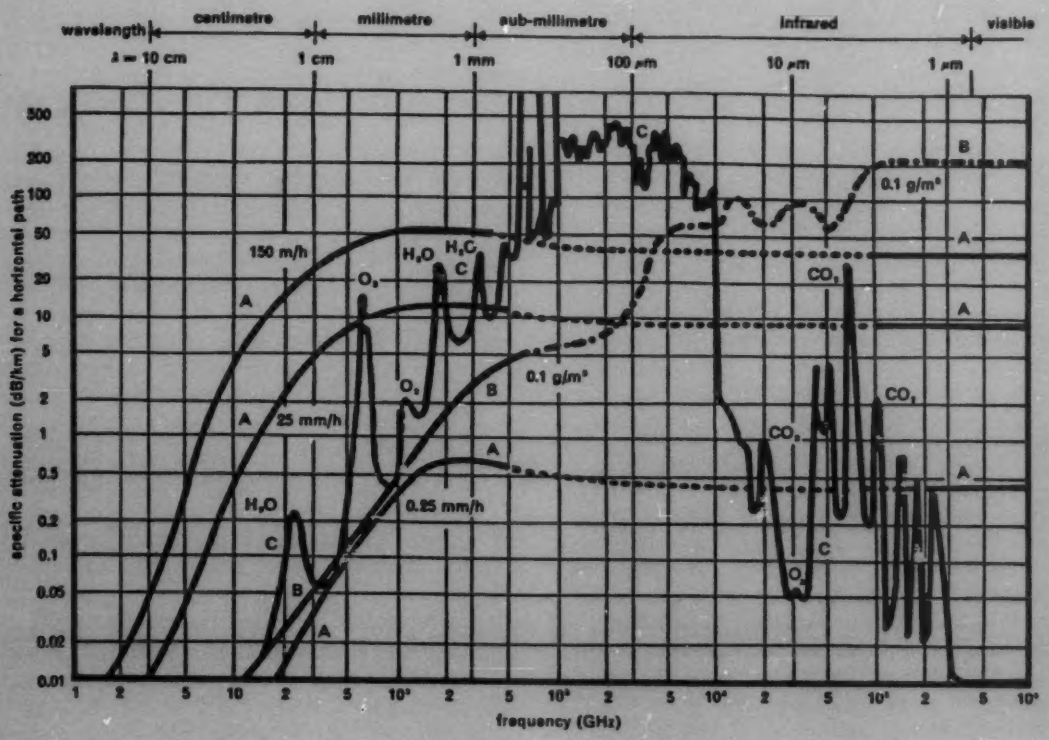
An approximate frequency scaling law for rainfall attenuation is:

$$\frac{A(f_2)}{A(f_2)} = \left(\frac{f_2}{f_1}\right)^{1.72} \qquad (5)$$

(which for uniform rain becomes approximately a_2/a_1).

It can be shown that water cloud attenuation is equivalent to light rain ($\leq$ 5 mm/Hr) attenuation below 100 GHz. Ice cloud attenuation is at least two orders of magnitude less and therefore of little relative practical significance.

Figure 3. Specific attenuation of radiowaves, IR, and visible light due to atmospheric constituents. The conditions assumed: T = 20°C, vapor pressure = 7.5 cm/m³. A = rain, B = fog, C = gas.



Polarization Effects in Earth-Space Paths

The Faraday effect is well known to ionosphericists since it is a major technique for measuring the total electron content of the ionosphere. The expression for the amount of Faraday rotation is given by:

$$\Omega = 2.97 \times 10^{-2} f^{-2} \int H \cos\theta \sec\chi \, Ndh \quad (6)$$

where Ω is the amount of rotation of the plane of polarization (radians), f is the radiofrequency (Hz), H is the magnetic field strength (ampere turns/meter), θ is the angle between the ray path and the geomagnetic field vector, χ is the ray zenith angle, N is the electron density (#/m^3) and h is the vertical distance.

Due to the inverse f^2 dependence of Ω, the amount of Faraday rotation diminishes rapidly as the space frequency is raised. At 7 GHz, for example, the plane of polarization of the transmitted signal would rotate only 1.4 degrees in transit through the entire ionosphere assuming a value for $\int Ndh$ of 10^{18} electrons/m^2 and H cos θ sec χ = 40 amperes turns/meter. Below 1 GHz, the amount of rotation becomes significant and circularly polarized antennas are employed to obviate the effect.

The major factors in the transmission path causing depolarization effects are hydrometeors, multi-path, and Faraday rotation. Of these three, the depolarization caused by rain dominates decisively at GHz frequencies. It has been shown experimentally that the polarization isolation is inversely proportional to the signal attenuation. Ice depolarization also occurs but is less significant.

Since Faraday rotation is negligible at EHF, it is not necessary to use circular polarization at GHz frequencies. Indeed, there may be good reason to advocate intentional use of linear polarization at EHF since depolarization may be less for linear than for circular polarization.

Propagation Delay in Earth-Space Propagation Paths

In general, the time required for a signal to traverse a distance s is given by:

$$T = \frac{1}{c} \int_0^s ds + \Delta T_t + \Delta T_i + T_1 + T_2 \quad (7)$$

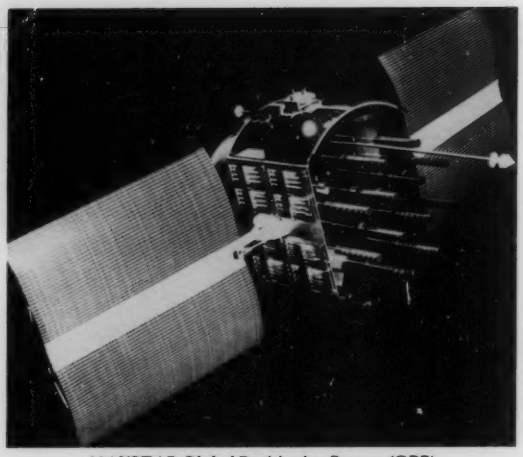
where $C^{-1} \int ds$ is the free space transit time, ΔT_t and ΔT_i are the delays due to propagation through the troposphere and ionosphere respectively, T_1 is due to pulse distortion arising from finite signal bandwidth, and T_2 is due to scattering.

The tropospheric and ionospheric delays can be written as:

$$\Delta T = \Delta T_t + \Delta T_i = \frac{10^{-6}}{c} \int_0^s N_t(s) ds + \frac{40.3}{cf^2} \int_0^s N_e ds \quad (8)$$

where c is the free space velocity of light, N_t is the tropospheric refractivity (n-1) x 10^6, f is the radiofrequency, N_e is the electron density, s is the distance along the ray trajectory, and the integral $\int_0^s N_e ds$ is the total electron content (TEC) of the ionosphere.

The tropospheric term ΔT_t is approximated by N_s H sec χ where N_s is the surface refractivity, H is the atmospheric scale height and χ is the ray zenith angle. Taking N_s = 300 and H = 7 X 10^3 meters, we see that ΔT_t is the order of 7 nanoseconds. This term may be easily modelled to within 1.5 nanoseconds. The ionospheric term ΔT_i is dependent upon frequency. At the Global Positioning System (GPS) frequency of 1.6 GHz, and taking the integral equal to 10^{18} electrons/m^2, we find that ΔT_i = 50 nanoseconds. There is, of course, a considerable variation in this number because of the extreme variability in the total electron content, TEC. Using various models for prediction of ΔT_i residuals from 1 and 13 nanoseconds have been observed.[37] Various models have been used to predict the TEC.[38,39,40,41,42] Of particular interest is an empirical model described by Klobuchar[38] that predicts the time delay for single frequency users of the NAVSTAR/GPS system. In general, the ionospheric contribution to time delay is compensated for through use of a two-frequency correction technique intrinsic to the GPS system. Thus, only single frequency users need to be concerned with the ionospheric contributions



NAVSTAR Global Positioning System (GPS)

to path delay. It will be seen shortly that by far the most important propagation effect for GPS is not propagation delay but scintillation.

In addition to a delay in the mean arrival time ΔT and the well-known pulse distortion effect T_1,[43] inhomogeneities in the ionospheric plasma give rise to a time delay spread of the radiowave signal (denoted by T_2). The time delay spread is directly related to angular scattering. The term T_1 arises from the fact that the various Fourier components of the signal travel at different speeds in the medium,[2] and the term T_2 is due to scattering from ionospheric turbulence.

Scintillation in Earth-Space Propagation

Fluctuations in signal power and phase often accompany radio wave propagation over earth-space paths as a result of inhomogeneties in the refractivity. This phenomenon, analogous to the twinkling of stars in the visible part of the electromagnetic spectrum, has been the object of research for several decades. Many excellent papers are available on ionospheric scintillation and we cannot do justice to this extremely rich and interesting phenomenon in this limited review. Nevertheless scintillation is probably the single most important deleterious factor affecting future systems utilizing the earth-space propagation path. Much experimental work has been conducted by Aarons and his coworkers at Air Force Geophysical Laboratory over the years and he has recently published a short summary of ionospheric scintillation that provides a good starting point for the uninitiated.[14] A thorough review has been presented by Crane.[44]

The morphology of ionospheric scintillation, which is of major interest here, is now fairly well established although details remain to be clarified. A considerable amount of effort has been directed toward the development of algorithms to describe the effect, with the ultimate objective of communication channel modelling. The approach has been to deduce the morphology from all available scintillation data and to derive the channel properties from the hypothesis of a two-component signal statistical model.[15] Alternate schemes for modelling the scintillation morphology based upon strong tendencies for correlation with Spread F[45,46] or upon the nature of the observed inhomogeneity wave number spectrum also have been suggested.[47,48]

Much of the current activity is focused on the scintillation cause–and–effect relationships, both in the auroral and the equatorial zones. However, more emphasis is placed on the latter zone where the effect is most intense. Indeed GHz scintillation over very limited regions may sometimes occur following ionospheric sunset near the geomagnetic equator. The most interesting aspects of the current drive to understand the problem stem from the involvement of three seemingly distinct phenomena; viz, radar backscatter of small scale structures, scintillation caused by ionospheric inhomogeneities of intermediate scale, and detection of quite large-scale electron content depletions (sometimes called bubbles or plumes). Clearly the instabilities that give rise to plume development are of major concern in understanding the equatorial scintillation problem. Moreover, the scintillation that occurs at high latitudes is thought to arise from an entirely different instability, and the modest scintillation observed at midlatitudes has not been fully investigated. Although plumes as such are not observed at high latitudes, large variations in TEC are at least circumstantially related to auroral zone scintillation since scintillation enhancements are conspicuous in the data only when the TEC gradient is exceedingly sharp.[49] A considerable advance in the

total understanding of ionospheric scintillation phenomenology as well as the underlying physical processes involved has been achieved through utilization of data sets (both amplitude and phase) obtained via the WIDEBAND DNA-002 program. It is well known that external factors related to sunspot activity strongly affect ionospheric scintillation occurrence and amplitude. Solar activity tends to enhance equatorial scintillation, and geomagnetic activity enhances scintillation near the auroral zone.

A considerable effort has been directed toward the elucidation of parameters that are important to the design of earth-space systems in order to counter the scintillation problem. Fluctuations in signal power are a major problem to satellite links in the military band (225-400 MHz) unless compensating techniques are implemented. Communication systems may counter the effects of substantial fading by using space diversity.

FLTSATCOM

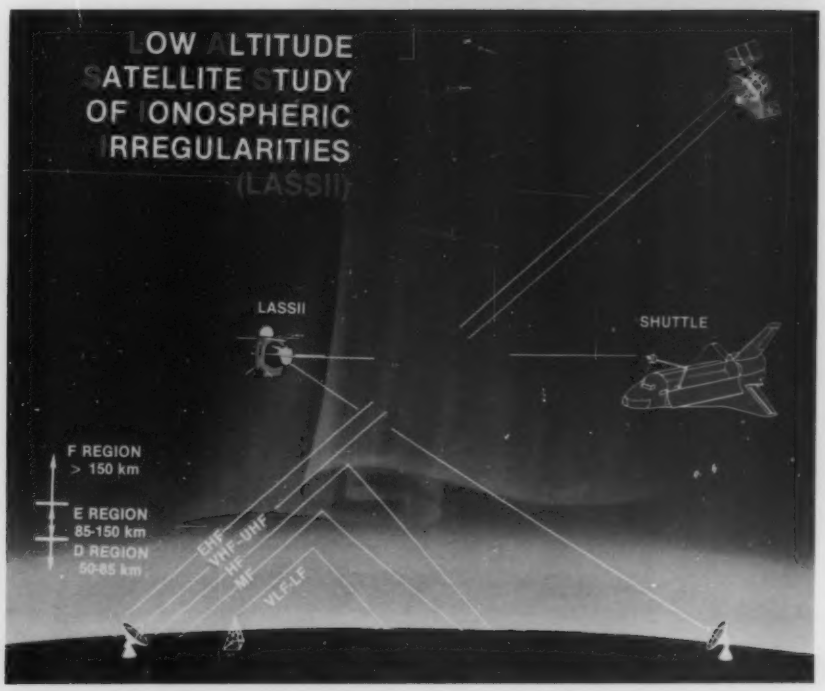


Figure 4.

The NRL LASSII Concept. Photo courtesy of E. P. Szuszczewicz, LASSII Program Manager

If the paths are sufficiently well separated (depending upon the details of the inhomogeneity wave number spectrum), then fading in the two links is uncorrelated and diversity gain may be achieved. Separations of the order of a kilometer are involved but these useful minimum separations are certainly larger than ship dimensions at UHF.[50] One would normally expect that radio links sufficiently separated in frequency, polarization, or transmission time would be effective in combating scintillation. Alas, this is not true in the case of polarization diversity, and furthermore frequency separation of up to 100 MHz may be required to obtain an adequate diversity gain. Clearly, frequency diversity is not applicable in the UHF band although it may be at higher frequencies where allocation problems are less severe. Consequently, time diversity is the only viable and demonstrated procedure for overcoming scintillation at UHF for disadvantaged mobile platforms. Coding and interleaving schemes have been investigated by Bucher,[51] White,[52] and Johnson.[53]

It is anticipated that the unprotected FLTBCST channel of FLTSATCOM will encounter significant disruption at solar maximum (1979-81) especially within ±20° of the geomagnetic equator. For platforms utilizing the AN/SSR-1 receiver and served by FLTSATCOM, it is anticipated that high latitude (i.e., Norwegian Sea) scintillation will distort message traffic in that zone up to 5% of the time, and equatorial (South and Central Pacific, South Atlantic, and Indian Ocean) scintillation will distort message traffic as much as 30% of the time following sunset and before midnight. However these are statistical results. Our current ability to predict the actual scintillation magnitude or rate at a particular place and time is virtually non-existent. Alternative mitigation or avoidance schemes besides brute force (increasing antenna gain or transmitter power on the uplink or downlink as appropriate) include utilization of DSCS assets at 7-8 GHz for FLTBCST. This involves the reception of UHF FLTBCST by specified gateway stations that are also equipped with DSCS terminals. These gateways would be located adjacent to but outside of virulent scintillation zones. Retransmission of the FLTBCST to users in the scintillation zone would be accomplished via DSCS at 7-8 GHz where scintillation is believed to be less severe. It should be noted, however, that selection of the gateway stations is rather a critical function of the known (presumed) morphology of scintillation, and that the absence of GHz scintillation in the equatorial zone is by no means clear.[54] As a matter of fact, significant scintillation is often observed at L-Band. Klobuchar[55] has monitored GPS satellite transmissions at both Kwajalein Atoll and Ascension Island and has detected scintillation of sufficient magnitude to suggest severe problems for users at those locations. Even though the elevated intensity of the equatorial scintillation at L-Band is undoubtedly related to the peak in solar activity, it nevertheless is of some concern.

The parameter most commonly used to describe the degree of scintillation encountered over an earth-space path is the so-called S_4 index. This index is easily approximated by scaling strip-chart recordings or through computation since S_4 is simply the normalized rms signal fluctuation. In practical terms, the most important parameters needed for propagation channel specification include S_4 (the scintillation index), T_c (the fade coherence time), and a rough measure of the coherence bandwidth. The fade coherence time must, of course, be large compared to the baud duration to avoid failure; nevertheless as T_c decreases the "time diversity coding gain" will typically increase until the baud duration limit is reached. In this regime, the S_4 index is found to be appropriate to the Nakagami[56] distribution which describes the probability distribution of amplitude scintillation adequately. As a result, fading level statistics can be deduced with a degree of confidence knowing S_4 alone and the system degradation introduced by S_4 can be diminished through time diversity. Of course there is some loss in timeliness and throughput efficiency. Clearly, in a Rayleigh fading environment (i.e. $S_4 = 1$) the scintillation is characterized as "strong" and pulse distortion may arise in some instances under the above conditions. It is interesting to note that a precise description of the channel using complex signal statistics (i.e. specification of both amplitude and phase) is unnecessary unless the fading is rapid and strong. Fortunately when this occurs, the Rayleigh model, which is well known and understood, may be employed. Thus, under intense scintillation conditions encountered over the geomagnetic equator, it is neces-

sary only to measure the fade coherence time T_c to determine the performance of a satellite link. This parameter is directly related to the speed with which the ionospheric inhomogeneities drift through the ray trajectory.

There is a growing interest in transmitting increasingly higher data rates through the earth-space path. As mentioned above, however, scintillation conditions in some instances may not support such a requirement. Furthermore, the need for greater accuracy and availability of precise navigation data serves to emphasize the constraints imposed by ionospheric inhomogeneities. Techniques for synthesizing large antennas in space by coherent processing and for improving the detection range of space surveillance radars by coherent detection rely heavily on ionospheric smoothness at the frequency involved. Thus, the temporal and spatial personality of ionospheric inhomogeneities and the associated phase and amplitude scintillations are of utmost relevance. The ionospheric limitation to coherent integration in trans-ionospheric radars has been discussed by Rino.[57] These authors find that the time variation of signal phase is given by

$$\phi(t) = \phi_o + w(t-t_o) + \dot{w}\,(t-t_o)^2 + \delta\phi(t-t_o) \qquad (9)$$

which is defined over a short interval ($t_o \leq t \leq t_o+T$) and $\delta\phi$ is a random component of phase defined by a power law probability density function.

The ϕ_o term is a phase bias arising from the total electron content along the ray path (TEC), the linear term $w(t-t_o)$ is the doppler shift (which causes no problem for target coherent detection), and $\dot{w}(t-t_o)^2$ gives rise to spectral broadening and may reduce processing gain. Processing gain is also diminished by the random term $\delta\ \phi\ (t-t_o)$. For midnight periods, both $\dot{w}$ and $\delta\ \phi$ are quite large over the equator and integration at VHF is limited to much less than 10 seconds. More striking is the fact that omnipresent non-vanishing $\dot{w}$ values during the day will limit integration to less than a minute.

Tropospheric scintillation is also observed in both clear air and cloudy environments. However the depth of fading is typically less pronounced than the ionospheric scintillation observed at lower frequencies. Ionospheric amplitude and phase scintillations diminish in proportion to $f^{-1.5}$ and f^{-1} respectively, but tropospheric scintillation exhibits little resolvable frequency dependence.[58]

Although both ionospheric and tropospheric scintillation increase as the zenith angle increases, the obliquity factor for the latter is much more severe. Near the horizon, tropospheric scintillation will dominate at most earth-space frequencies except under the more virulent ionospheric conditions. The variance of scintillation (scintillation index) is found to be proportional to $L^{11/6}$ where L is the tropospheric path length. Scintillation as high as 25 dB has been observed at elevation angles of 2° in the 20-30 GHz band.[58]

Doppler Frequency and the Earth-Space Path

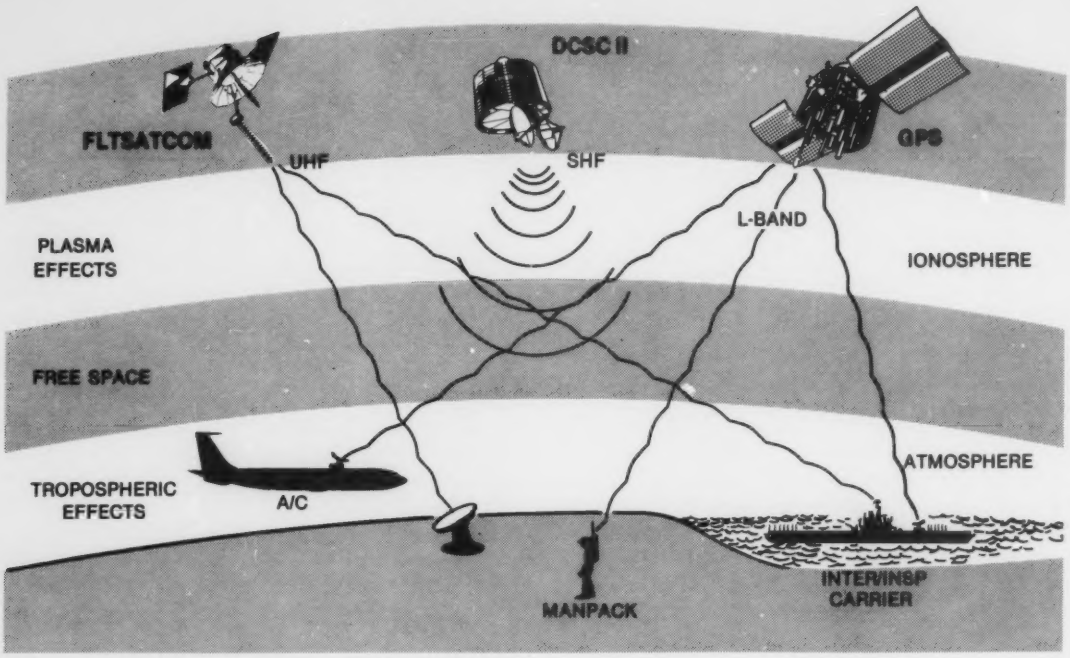
The well known expression for excess doppler (Hz) introduced by the ionosphere is:

$$\Delta f = -\frac{40 \times 10^6}{cf} \frac{d}{dt} \int N_e ds \qquad (10)$$

where f is the transmission radio frequency, c is the speed of light, and N_e is the electron density.

This number is typically negligible in comparison with the free-space doppler introduced by satellite or target motion.

For geosynchronous satellites, the Δ f correction to the transmission frequency is only academic (being only a fraction of a Hertz), and even for orbiting satellites it is relatively unimportant at typical space frequencies. The maximum value of the time derivative occurs near the horizon. Even in this extreme case and for $\int N_e ds = 10^{18}$ electrons/m^2, Δf will be less than 5 Hz at 1.6 GHz. Well-designed ϕ-lock tracking loops will encounter no difficulty. Even though the dispersive doppler introduced by the ionosphere is not significant as a system



effect in earth-space propagation, it has been quite useful in ionospheric studies.

A Commitment to Research

There are numerous programs underway at NRL to elucidate the intrinsic properties of the troposphere, ionosphere, magnetosphere, and beyond. The realm of solar physics and the myriad of solar-terrestrial relationships are fundamental concerns and the Navy was at the cutting edge in developing concepts for monitoring solar electromagnetic and corpuscular flux. The SOLRAD program, for example, served as a basis for many of the current solar monitoring systems that serve the DoD and the scientific community. Studies of magnetospheric dynamics are underway and there is major interest in developing an understanding of both the benign and disturbed ionosphere. The proposed NRL LASSII program–an acronym standing for *L*ow *A*ltitude *S*atellite *S*tudy of *I*onospheric *I*rregularities–has been developed to address many of the unknowns that still exist in the near-earth plasma. LASSII is a sortie-mode satellite designed to exploit the Space Shuttle in studies of natural and artificially-induced irregularities and their associated effects on radiowave systems used by the military.[59] As depicted in Figure 4, the program will address not only earth-space propagation but also systems that employ frequencies which interact with the lower ionosphere most dramatically (i.e., HF and below). The program will focus on short-duration (7-30 days) experiments that will take advantage of extensive built-in flexibility in the LASSII payload complement. The total program is an inter-disciplinary one involving not only basic physics but also addresses the myriad of practical problems that concern system designers who must deal with the complex and incompletely understood radiowave propagation environment. The experiment duration and the plan for multiple missions will accommodate the study of numerous concepts in geoplasma assessment, modification, and control, and will provide the basis for near-term solutions to important research and development problems.

Earth-space propagation effects are currently being studied through utilization of radio receivers placed in the more virulent regions of the globe. Of primary interest are studies of effects on UHF channels of FLTSATCOM, on the L-band channels of GPS, and on future EHF systems. A major investigation of deleterious effects is especially necessary in the maritime environment, where archival data are sparse, and NRL is launching a campaign to collect pertinent information during the period of maximum solar activity (1980-1982). The areas of primary interest include equatorial zones near Ascension, Diego Garcia, and Guam and high latitude zones near the Norwegian and North Seas.

Conclusion

There are numerous applications for utilization of space in the arenas of communication, navigation, surveillance and related disciplines. The one unique advantage afforded by space is the vantage point it provides. A single satellite, appropriately placed in geosynchronous orbit, can observe and/or serve almost 1/2 the globe. The trend in DoD is for satellite platforms to be the backbone for most strategic and tactical communications, navigation, positioning and overhead surveillance. The virtues of satellites for use in commercial communication and remote sensing of earth resources, weather systems, ocean environment and related areas are well known. Furthermore the use of satellites to relay data from earth terminals, buoys, and other satellites is of major significance.

The trend toward increased utilization of space must be tempered by a parallel awareness that overzealous commitment could ultimately be an Achilles heel. Factors such as survivability in a strategic environment are being studied since they are obvious considerations. However, system architects should also be aware of the environmental constraints which the use of space will necessarily introduce. The ionospheric parameters of importance in space system design are known, but their detailed personalities are not completely understood, and forecasting capability is almost non-existent. Typically, ad-hoc climatologies are employed to define the ranges (i.e. margins) over which systems must be made to adapt. Thus ionospheric and tropospheric research has been of great benefit to system designers in specifying, to first order, the degree of robustness that must be engineered into space systems. Design constraints would be further reduced through use of second-order improvements involving some form of environmental monitoring or assessment function. Increasingly, the point of view is emerging that the short-term forecasting requirement must be achieved through quasi-real-time environmental remote sensing that is employed in conjunction with algorithms for extrapolation into denied areas.

Not to be ignored is the consideration that environmental limitations to earth-space propagation apply to adversary as well as friendly forces. It is not inconceivable that techniques for exploitation and/or control of the environment will be components in the hierarchy of future electronic warfare systems. Of special interest in this regard are emerging studies of ionospheric modification using chemical reagents and RF heating. Natural disturbances also have morphologies that might also be exploited, although current capability to forecast natural events is probably insufficient. With the increased demands placed upon space systems in terms of accuracy and data rate, the environmental constraints upon the earth-space path may well be a limiting factor in the ability to achieve the design goals. The successful search for "windows" or "doors" in which environmental constraints exhibit extremes should enable enhanced operation in both natural and disturbed environments and more secure operation against a threat. To the extent that "doors" and "windows" are non-existent in nature, their creation may be critical and more than justifies continued research in the arena of environmental modification. ■

References

1. Schemmer, B. F., 1978, "Strategic C^3: The Satellite Arena–20 Years of the Sputnik," Armed Forces Journal International, February. pp. 18-30.

2. Millman, G. H., 1967, "A Survey of Tropospheric, Ionospheric, and Extra-Terrestrial Effects on Radio Propagation Between the Earth and Space Vehicles," in *Propagation Factors in Space Communications*, edited by W. T. Blackband, NATO-AGARD Conference Proc., Technivision, Maidenhead, England; also G. E. Report TISR66EMHI, pub. 1965.

3. Lawrence, R. S., C. G. Little, and H. J. A. Chivers, 1964, "A Survey of Ionospheric Effects Upon Earth-Space Radio Propagation," Proc. IEEE, Vol. 52, pp. 4-27.

4. Davies, K., 1965, *Ionospheric Radio Propagation*, NBS Monograph 80, Supt. Doc., U.S. Gov't Printing Office, Washington, D.C.

5. Kelso, J. M., 1964, *Radio Ray Propagation in the Ionosphere*, McGraw-Hill, New York.

6. Budden, K. G., 1964, *Lectures on Magnetionic Theory*, Gordon and Breach, New York.

7. Kerr, D. E., 1951, *Propagation of Short Radio Waves*, McGraw-Hill, New York (Also Dover Publications, New York, 1965).

8. Battan, L. J., 1959, *Radar Meteorology*, Univ. of Chicago Press, Chicago.

9. Bean, B. R., and E. J. Dutton, 1966, *Radio Meteorology*, NBS Monograph 92, Supt. Doc., GPO, Washington, D.C.

10. Ippolito, L. J., 1975, Unpublished lecture notes for GWU Course 501, *Millimeter Wave Space Communications.*

11. Rush, C. M., 1979, "Transionospheric Radio Propagation", in *Aerospace Propagation Media Modelling and Prediction Schemes for Modern Communications Navigation, and Surveillance Systems,* edited by H. Soicher, AGARD-LS-99, Tech. Edit. and Reprod. Ltd., London.

12. Klobuchar, J. A., 1977, "Ionosphere Time Delay Corrections for Advanced Satellite Ranging Systems," in *Propagation Limitations of Navigation and Positioning Systems*, AGARD-CP-209, Tech. Edit. and Reprod. Ltd., London.

13. Klobuchar, J. A., 1978, "Ionospheric Effects on Satellite Navigation and Air Traffic Control Systems," in *Recent Advances in Radio and Optical Propagation for Modern Communications, Navigation and Detection Systems*, edited by J. Aarons, AGARD-LS-93, Tech. Edit. and Reprod. Ltd., London.

14. Aarons, J., 1978, "Ionospheric Scintillations—An Introduction," in *Recent Advances in Radio and Optical Propagation for Modern Communications, Navigation and Detection Systems*, edited by J. Aarons, AGARD-LS-93, Tech. Edit. and Reprod. Ltd., London.

15. Fremouw, E. J. and C. L. Rino, 1978, "A Signal Statistical and Morphological Model of Ionospheric Scintillation," in *Operational Modelling of the Aerospace Propagation Environment*, edited by H. Soicher, AGARD-CP-238, Tech. Edit. and Reprod. Ltd., London.

16. Crane, R. K., 1977, "Prediction of the Effects of Rain on Satellite Communication Systems," Proc. IEEE, Vol. 65, #3.

17. Bean , B. R. and E. J. Dutton, 1976, "Radio Meteorological Parameters and Climatology," *Telecomm. J.*, 43 (VI), 427-435.

18. Boithias, L., 1976, "Structure of the Tropospheric Refractive Index and Propagation," *Telecomm. J.*, 43 (VI), 419-426.

19. CCIR Report 263-3, 1974, "Ionospheric Effects on Earth-Space Propagation."

20. Goodman, J. M., (editor), 1975, "*Effect of the Ionosphere on Space Systems and Communication*," GPO, Washington, D.C.

21. Goodman, J. M., (editor), 1978, "*Effect of the Ionosphere on Space and Terrestrial Systems*," GPO, Washington, D.C.

22. Mendillo, M., (editor), 1976, "The Geophysical Use of Satellite Beacon Observations," Boston University Pub.

23. Bean, B. R. and R. E. McGavin, 1965, "A Review of Refraction Effects on the Apparent Angle of Arrival of Radio Signals," in *Propagation Factors in Space Communications*, edited by W. T. Blackband, NATO-AGARD Conf. Proc., Technivision, Maidenhead, England.

24. Bean, B. R. and B. A. Cahoon, 1957, "Use of Surface Weather Observations to Predict the Total Atmospheric Bending of Radiowaves at Small Angles," Proc. IRE 45, 145-6.

25. Bean, B. R., G. D. Thayer, and B. A. Hart, 1971, "Worldwide Characteristics of Refractive Index and Climatological Effects," in *Tropospheric Radio Propagation (Part 1)*, edited by H. J. Albrecht, NATO-AGARD-CP-70; Tech. Edit. and Reprod. Ltd., London.

26. Albrecht, H. J. (editor), 1971, *Tropospheric Radio Propagation (Parts 1 & 2)*, NATO-AGARD-CP-70-71, Tech. Edit. and Reprod. Ltd., London.

27. Millman, G. H., 1971, "Tropospheric Effects on Space Communications", in *Tropospheric Radio Propagation (Part 1)*, edited by J. J. Albrecht, NATO-AGARD-CP-70, Tech. Edit. and Reprod. Ltd., London.

28. Bean, B. R., B. A. Cahoon, C. A. Samson, and G. D. Thayer, 1966, "A Worldwide Atlas of Radio Refractive Index," ESSA Manograph 1, U.S. GPO, Washington, D.C.

29. Gerson N. C., (editor), 1962, *Radiowave Absorption in the Ionosphere*, AGARDOGRAPH 53, Pergamon Press, New York.

30. Kaul, R., R. Wallace, and G. Kimal, "A Propagation Effects Handbook for Satellite Systems Design", NASA Rpt. ORI TR 10.

31. Laws, J. O. and D. A. Parsons, 1943, "The Relation of Raindrop Size to Intensity," Trans. AGU, 24, 452-460.

32. Marshall, J. S. and W. M. Palmer, 1948, "The Distribution of Raindrops with Size," *J. Meteorology* 5, 165-166.

33. Joss, J., J. C. Thams, and A. Waldvogel, 1968, "The Variation of Raindrop Size Distributions at Locarno," Proc. Int. Conf. Cloud Physics, Toronto, pp. 369-373.

34. Ryde, J. W. and D. Ryde, 1945, "Attenuation of Centimeter and Millimeter Waves by Rain, Hail, Fog and Clouds", Report 8670, G. E. Research Labs., Wembley, G. B.

35. Gunn, K. L. S. and T. W. R. East, 1954, "The Microwave Properties of Precipitation Particles", Quarterly J. Roy. Met. Soc., 80, 522-545.

36. Beach, J. B., 1979, "Atmospheric Effects on Radiowave Propagation," *Defense Electronics*, Dec., pp. 95-98.

37. Parkinson, B. W., E. M. Lassiter, and C. K. Cretcher, 1977, "Ionospheric Effects in NAVSTAR/GPS," in *Propagation Limitations of Navigation and Positioning Systems*, edited by P. Halley, AGARD-CP-209, Tech. Edit. and Reprod. Ltd., London.

38. Klobuchar J. and R. Allen, 1970, "A First-Order Prediction Model of Total Electron Content for a Mid-latitude Ionosphere," AFCRL-70-0403.

39. Waldman, H. and A. V. DaRosa, 1971, "Prognostication of Ionospheric Electron Content," SAMSO TR-71-82, DDC-AD 731095/6, Stanford University.

40. Rao, N. N., M. Y. Yoreakim, and K. C. Yeh, 1971, "Feasibility Study of Correcting for the Excess Time Delay of Transionospheric Navigational Ranging Signals," SAMSO TR-71-63, DDC-AD-729797, University of Illinois.

41. Pisacane, V. L., M. M. Feen, and M. Sturmanis, 1972, "Prediction Techniques for the Effect of the Ionosphere on Pseudo-Ranging from Synchronous Altitude Satellites," SAMSO TR-72-22, DDC-AD 749486, APL.

42. Bent, R. B., S. K. Llewellyn, and M. K. Wallock, 1972, "Description and Evaluation of the BENT Ionospheric Model," SAMSO TR-72-239, DDC-AD 753081-6, DBA Systems, Melbourne, Fla.

43. Wong, Y. K., K. C. Yeh, and C. H. Liu, 1978, "Mean Arrival Time and Mean Pulse Width of Signals Propagating Through an Inhomogeneous Ionosphere with Random Irregularities," in *Effect of the Ionosphere on Space and Terrestrial Systems*, edited by J. M. Goodman, GPO, Washington, D.C.

44. Crane, R. K., 1974, "Morphology of Ionospheric Scintillation," Lincoln Lab., TN-1974-29.

45. Singleton, D. G., 1979a, "Predicting Transionospheric Propagation Conditions," paper prepared for *STP-PW Conference*, Boulder.

46. Singleton, D. G., 1979b, "An Improved Ionospheric Irregularity Model," paper 59 prepared for *STP-PW Conference*, Boulder.

47. Basu, S., S. Basu, and B. K. Khan, 1976, "Model of Equatorial Scintillation from In-Situ Measurements," *Radio Sci.*, 11, 821.

48. Basu, S. and S. Basu, 1979, "Model of Phase and Amplitude Scintillations from In-Situ Data," paper 129 prepared for *STP-PW Conference*, Boulder.

49. Rino, C. L., 1979, "Some Ramifications of the Wideband Satellite Data for Scintillation Modelling," paper prepared for *STP-PW Conference*, Boulder.

50. Paulson, M. R. and R. U. F. Hopkins, 1973, "Effects of Equatorial Scintillation Fading on SATCOM Signals," NELC/TR/1875.

51. Bucher, E. A., 1975, "UHF Satellite Communications During Scintillation," Lincoln Lab. TN 1975-10.

52. White, D. P., 1977, "A Time Diversity Coding Experiment for a UHF/VHF Satellite Channel with Scintillation: Equipment Description," Lincoln Lab. TN 1977-22.

53. Johnson, A. L., 1975, "Simulation and Implementation of a Modulation System for Overcoming Ionospheric Scintillation Fading," in *Radio Systems and the Ionosphere*, edited by W. T. Blackband, AGARD-CP-173, Tech. Edit. and Reprod. Ltd., London.

54. Craft, H. D. and L. H. Westerlund, 1972, "Scintillations at 4 and 6 GHz Caused by the Ionosphere," paper presented at AIAA 10th Aerospace Sciences Meeting, San Diego.

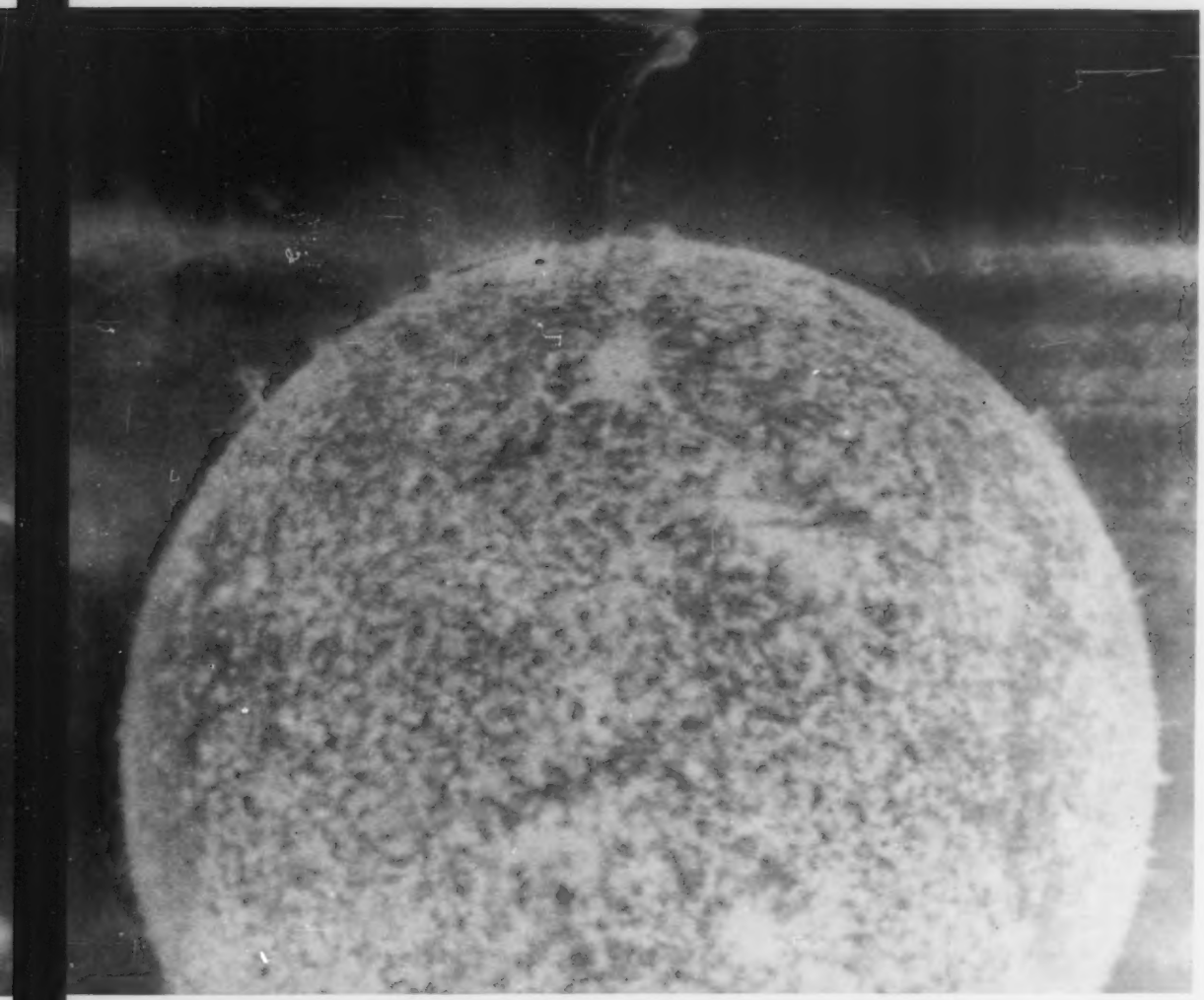
55. Klobuchar, J. A., 1980, private communication.

56. Nakagami, M., 1960, "The M-Distribution–A General Formula for Intensity Distribution of Rapid Fading," in *Statistical Methods on Radio Wave Propagation*, edited by W. C. Hoffman, Pergamon Press, New York.

57. Rino, C. L., C. H. Hawson, R. G. Livingston, and J. Petriceks, 1978, "The Ionospheric Limitation to Coherent Integration in Transionospheric Radars," in *Effect of the Ionosphere on Space and Terrestrial Systems*, edited by J. M. Goodman, GPO, Washington, D.C.

58. Hodge, D. B., D. M. Theobold, and R. C. Taylor, 1976, "ATS-6 Millimeter Wavelength Propagation Experiment," Ohio State Univ. Electro Science Lab., Report 3863-6.

59. Szuszczewicz, E. P., 1980, "NRL-701/LASSII," Memorandum Report 4149.



Study of Iron in Magnetotactic Bacteria

by

Richard B. Frankel
Massachusetts Institute of Technology
Richard P. Blakemore
University of New Hampshire, Durham

The authors would like to acknowledge the collaboration of Drs. Adrianus Kalmijn and Charles Denham of Woods Hole Oceanographic Institution on some of the research.

Bacteria have recently been discovered directing their movements by sensing the earth's magnetic field. They contain arrays of crystals of lodestone or magnetite which they synthesize from iron in the ocean. The discovery of the same type of crystals in the abdomens of honey bees and the brains of homing pigeons suggests that nature is not oblivious to the earth's magnetic field. The biological mechanism for sensing magnetic fields has never been determined. These bacteria are capable of biologically synthesizing highly purified single domain magnets that would be very expensive to produce chemically.

The tiny perfect magnetite crystals made by the bacteria possibly could be utilized in microcircuitry. In addition, because these bacteria are attracted to metal surfaces surrounded by distortions in the earth's magnetic field, they may play a role in ship fouling. The unique physical properties of the single domain ferromagnetic crystals of magnetite produced by these bacteria are being analysed for possible electronic applications. Biochemistry of heavy metal storage could be important in recovering metals from ore and nuclear wastes; this is a continuing research project supported by the Office of Naval Research.

Several species of aquatic bacteria which orient in the earth's magnetic field and swim along magnetic field lines in a preferred direction (magnetotaxis) have been observed in marine and freshwater sediments of the northern hemisphere. Their orientation is due to one or more intracytoplasmic (intracellular) chains of single-domain magnetite particles. These linearly arranged particles impart a net magnetic dipole moment to the bacterium, parallel to the axis of motility. Northern hemisphere magnetotactic bacteria with unidirectional motility swim consistently in the direction of the magnetic field, that is, to the geomagnetic north. This implies that their magnetic dipole is systematically oriented with the north-seeking pole forward. Bacteria from aquatic environments in New Zealand and Australia orient in the earth's magnetic field and when separated from the substrate (bottom) swim along magnetic field lines to the south. This implies that their magnetic dipole is oriented with the south-seeking pole forward. Consequently, both northern- and southern-hemisphere magnetotactic bacteria observed to date migrate downward by swimming along the earth's inclined magnetic field lines. (Figure 1)

Magnetotactic bacteria were first discovered in sediments near Woods Hole, Massachusetts in 1974 by R. P. Blakemore. It was observed that most of these bacteria orient and swim along the magnetic lines toward the North. Reversal of the ambient magnetic field by Helm-

Figure 1: *Negative stain of a magnetic bacterium collected from a fresh water pond. This bacterium was magnified 50,000 times with a transmissiom electron microscope.*
Bar = 1 micrometer

holtz coils caused the cells to make U-turns within one second and swim in the opposite direction. Killed cells also oriented in uniform fields as low as 0.1 G. In these and other respects, the cells behave like single magnetic dipoles.

A magnetotatic spirillum, designated strain MS-1, has been isolated from a freshwater swamp and grown in pure culture (Figure 2). Each magnetotatic cell has an intracytoplasmic chain of approximately 22 iron-rich, electron-opaque particles. Each particle is roughly octahedral, 500 Å along each major axis. Mössbauer spectroscopy of ^{57}Fe in freeze-dried cells shows conclusively that the particles are primarily composed of magnetite, Fe_3O_4. Another iron-containing material with a room temperature Mössbauer spectrum similar to that of the iron-storage protein ferritin is also observed. Thus magnetotaxis is associated with intracellular magnetite. Since the bacteria were grown in chemically-defined media containing soluble (chelated) iron, the

Dr. Richard B. Frankel is Senior Research Scientist at the Francis Bitter National Magnet Laboratory, Massachusetts Institute of Technology. His research is in the areas of iron in biological systems, biomagnetism, Mössbauer spectroscopy, and biological effects of high magnetic fields.

Professor R. D. Blakemore is with the Department of Microbiology, University of New Hampshire, Durham. His research in the field of microbial physiology and diversity is currently emphasizing magnetotaxis in bacteria.

Figure 2. *Negative stain of magnetic spirillum strain 1. This bacterium was magnified 50,000 times with a transmission electron microscope. Bar = 1 micrometer*

presence of intracellular magnetite implies a process of bacterial synthesis.

The magnetic properties of magnetite particles depend on their size and shape. For a particle of roughly cubic shape with side dimension d, there is a range of d over which the particle will be a single magnetic domain. The magnetic moment of a single domain is the saturation or maximum magnetic moment of the particle and is unchanging in time. Magnetite particles with dimensions $d \leq 400$ Å are superparamagnetic, that is, thermal activation induces transitions of the magnetic moment between different easy magnetic axis directions with a consequent loss of magnetic "memory." Particles with dimensions $d \geq 800$ Å are multidomain and consequently have macroscopic moments which are either zero or less per unit volume than single domain particles. With d = 500 Å, the magnetite particles in strain MS-1 are within the single domain size range.

Because of strong interparticle interactions, the preferred orientation of the individual particles is such that their axes of magnetization are parallel, north-to-south along the chain direction. Thus, the entire chain acts as a single magnetic dipole with a moment equal to the sum of the particle moments. For a cell containing the average chain length of twenty-two particles, the total moment $M = 1.3 \times 10^{-12}$ emu. In the geomagnetic field of 0.5 G, the total magnetic energy of a cell $MH = 6.6 \times 10^{-13}$ erg. This value is more than an order of magnitude greater than thermal energy, kT (4.1×10^{-14} erg at 300 K). Thus each bacterium contains a sufficient, but not excessive, amount of single-domain-sized magnetite in an appropriate configuration to be oriented in the earth's magnetic field at ambient temperature, i.e., the cell's chain of magnetite crystals functions as a biomagnetic compass.

The simplest hypothesis for the mechanism of magnetotaxis is passive orientation of the bacterium resulting from the torque exerted by the ambient magnetic field in which it swims. Since most magnetotactic bacteria from the northern hemisphere are observed to swim northward, the compass in these cells must have a fixed orientation with respect to the flagellum (tail), with the north-seeking pole opposite to the flagellum. This orientation could be preserved in cell division if the com-

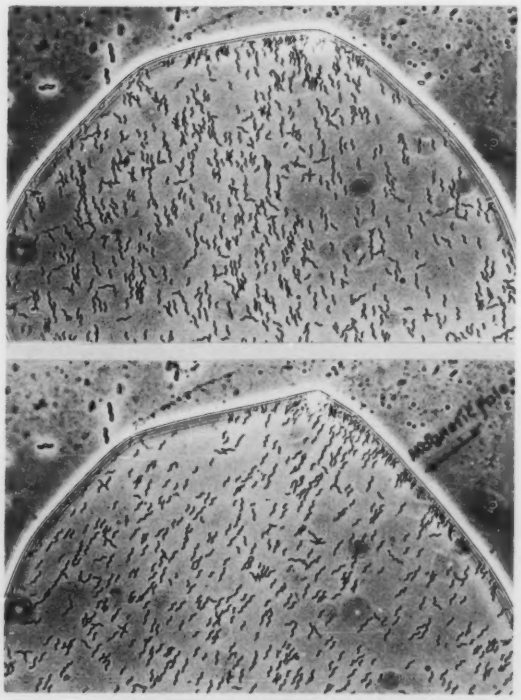


Figure 3. *Phase-Contrast Microphotograph showing alignment of swimming cells of strain MS-1 in an artificial magnetic field. Note accumulation of cells on the side nearest the magnetic pole. Magnified 330 X.*

pass is partitioned between the two daughter cells. Subsequently, during magnetite biosynthesis the magnetic moments of nascent magnetite particles at the ends of the pre-existing chains would become oriented along the chain direction by interaction with the chain dipole moment.

Due to the inclination of the earth's magnetic field, magnetotactic bacteria that swim to the north in the northern hemisphere are directed downward at an angle increasing with latitude. It has been suggested that this downward directed motion confers a biological advantage by guiding the bacteria, when dislodged, back to the sediments. On the basis of this hypothesis, magnetotactic bacteria of the southern hemisphere would be expected to swim south in order to reach the bottom. Recently, several morphological types of magnetotactic bacteria have been observed in sediments of Australia and New Zealand. These bacteria indeed swim consistently to the south, hence downward along the earth's inclined magnetic field lines, as hypothesized. As revealed by electron microscopy, they contain internal chains of electron-opaque particles similar to those observed in magnetotactic bacteria from the northern hemisphere. Like their northern-hemisphere counterparts, their magnetic polarity can be permanently reversed and they cannot be demagnetized.

The prevalence of south-seeking magnetotactic bacteria in southern-hemisphere sediments and north-seeking magnetotactic bacteria in the northern hemisphere verifies the hypothesis that downward directed motion is advantageous for, and upward directed motion detrimental to, the survival of these magnetotactic bacteria with unidirectional motility. Magnetotaxis is a reliable means of keeping these microorganisms in or near the bottom sediments. Since particles in the micrometer-size range with densities close to 1.0 tend to remain suspended in water, gravity is virtually inconsequential in determining the vertical distribution of the bacteria. ■

Bibliography

1. Frankel, R. B. and Blakemore, R. P. (1980). Navigational Compass in Magnetic Bacteria. *J. Magn and Magn Matls. 15-18*:1562-1564.

2. Blakemore, R. P., Frankel, R. B. and Kalmijn, Ad. J. South-Seeking Magnetotactic Bacteria in the Southern Hemisphere
Nature (in press)

3. Denham, C. R., Blakemore, R. P. and Frankel, R. B. (1980). Bulk Magnetic Properties of Magnetotactic Bacteria *IEEE Trasns.* on magnetics (in press).

The Acoustic Microscope: A Concept for Microscopy Using Waves of Sound

by

C. F. Quate
Stanford University

Professor Quate is Chairman of the Applied Physics Department at Stanford. He is a member of the National Academy of Engineering and the National Academy of Sciences. His research and many publications have made him reknowned in the field of acoustic imaging.

Microwaves–Sound Waves–Cryogenic Fluids: three different areas in the physical world. Each is important, and each is a traditional area of research for the Office of Naval Research. By combining the technologies in these three areas, something is achieved that is quite new, and important for the technological base of the U.S. Navy. These combined technologies can produce a new system for viewing the microscopic world and exploring a host of microscopic structures. In that region beyond the reach of the human eye, detailed information on a variety of diverse objects is needed.

- Steel is used in the form of huge plates familiar to every shipyard. Its important properties (strength, ductility, deformability) depend on the grains that make up the entire mass. Each grain, microscopic in size, has individual properties that must be understood. How does it adhere to its neighbor, and how is it oriented?
- Integrated circuits as fabricated on wafers of silicon are important elements of many electronic systems used by the Navy. Even though these structures are microscopic in size, there is a research goal to reduce their dimensions even further. It is important to develop systems for monitoring these fabricated structures.
- Biological cells, the basic entities that determine the well-being and health of naval personnel, are natural microscopic objects, and new methods for viewing may increase our understanding and knowledge in that area.

The extension of human vision has long been important to the Navy. Devices as common as binoculars and as complex as radar systems are used to extend our vision. Today satellites with high quality photographic equipment installed on board permit us to "see" the entire world.

Microscopes extend our vision into the region of the ultrasmall, which encompasses most of the material that makes up our physical world. The elastic properties of materials can now be examined in minute detail with a new microscope – a microscope based on sound waves. These new instruments have a resolving power that rivals the optical microscope. This article describes an ONR supported research program that proposes building a sound microscope with a resolution that surpasses what we now have.

Microwaves, the common form as found in radar systems, are not suitable for microscopes. Since the wavelength at 1500 MHz is 20 cm, one cannot expect to resolve objects smaller than that. Microscopes are not needed, for that is a job well-suited for the unaided human eye. However, microwaves are superb for radar and telescopes. Modern radioastronomers now have antennas for listening to the emission of these waves from regions of outer space.

By transforming electromagnetic energy at 1500 MHz into sound waves, which can be done with a piezoelectric film, the wavelength becomes smaller. It is the velocity of sound that makes the difference. The sound velocity in water is less than the velocity of electromagnetic waves by a factor of 200,000. The wavelength is reduced by the same factor, and the 20 cm length is reduced to 1 micron. In liquid helium, a cryogenic fluid, the wavelength becomes even smaller, 1/6 micron. A proper imaging system should allow us to resolve details of that size.

Microscopes, whatever their variety, have two features in common. The first is that they both use radiation with a wavelength that is comparable to the smallest objects being observed. The wavelength of visible light is 0.5 micrometers and this number establishes the optical resolving power. In the electron microscope the ultimate limit of resolution is determined by the wavelength of the electron packet. This allows one to view objects that are 10 to 100 times smaller than those found in optical micrographs. Radiation in the form of X-rays has a wavelength that is similar to the distance between atoms in a crystal lattice. An X-ray microscope should produce magnificent images.

The second factor common to microscopes is the lens. Optical lenses suitable for scientific instruments are quite complex with multiple surfaces cemented together to form a single unit. Lenses suitable for focusing X-rays do not exist and this alone prevents us from realizing the full potential of X-ray microscopy. It is our good fortune that lenses for acoustic waves are simple and easy to fabricate.

The Concept

Sound at microwave frequencies in liquid gives us wavelengths short enough for the microscope. A suitable lens will focus the radiation into a well defined waist. Such a lens serves to convert a plane wave of sound into a spherically converging wave. The energy is brought into sharp focus with a beam diameter that is less than one wavelength. This type of focusing where the beam diameter is controlled by wave diffraction can be achieved if the lens is free from aberrations. With optical lenses the problem is severe, but lenses have evolved over the years which overcome the aberrations. The acoustic lens is simpler, a single spherical surface between a high velocity material such as sapphire and a low velocity liquid such as water will suffice.

The action of a lens depends on the laws of refraction, the bending of rays as they pass from one medium to another. Even though the laws were worked out originally for optical waves, they apply equally to sound waves. Figure 1 illustrates the relationships. The angle of refraction according to Snell's Law increases as the veloc-

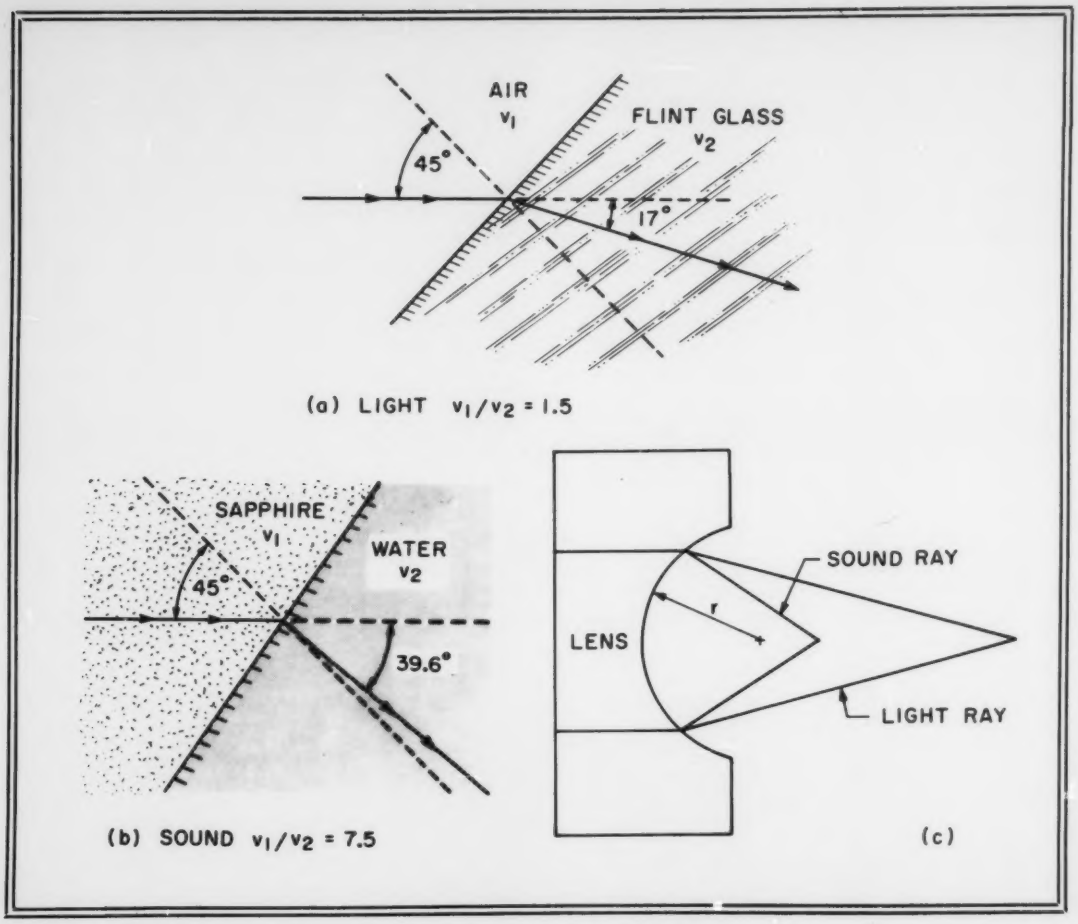


Figure 1: *The application of Snell's Law for light waves (1a) and sound waves (1b). A spherical surface (1c) will have a different focal length for these two wave types.*

ity ratio between the two media increases. For optical waves the velocity ratio might be 1.5, a rather modest value. For sound waves the velocity ratio is much larger. The velocity in sapphire is 7.5 times the velocity in water and 44 times the velocity in liquid helium. This makes an enormous difference in the focusing power of simple lenses. The comparison is made in Figure 2.

The remainder of the system, as sketched in Figure 3, is not unlike other systems in use throughout the Navy. It can be compared to a microwave radar system where a series of pulses is radiated and where the reflected echoes are analyzed for their information content or to sonar in miniature.

The Lens

At Stanford the initial work in the field of microwave acoustics was carried out under ONR's part of the Joint Services Electronics Program. That program with its long term continuous support has produced a number of striking advances in the field of electronics, such as delay lines, filters and signal correlators, all based on acoustic waves. The piezoelectric film transducer came from that program. It is very efficient as a converter of electromagnetic energy into acoustic energy and an important element in our imaging system.

As for the focusing system, the lens is ideal in the sense that it is free from aberrations. It serves to condense the energy into a region that measures less than one

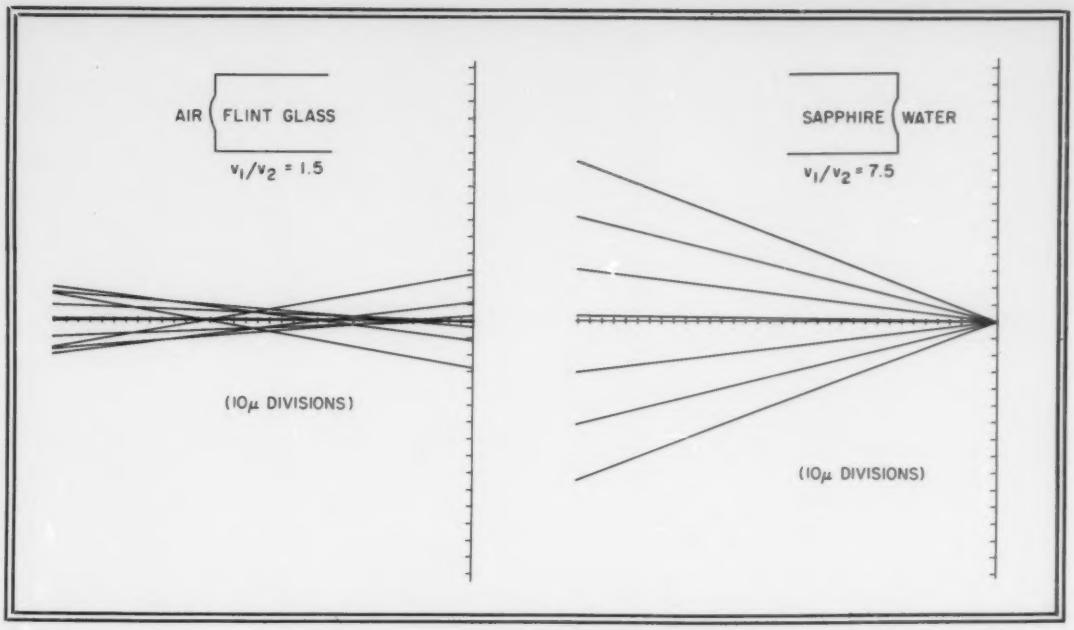


Figure 2: *A comparison of the focusing strength of the optical system (left) and the acoustic system (right) as illustrated with ray tracing. The weak refraction of optical waves produces spherical aberrations which limit the minimum size at the beam waist. The aberrations in the acoustic case are not important.*

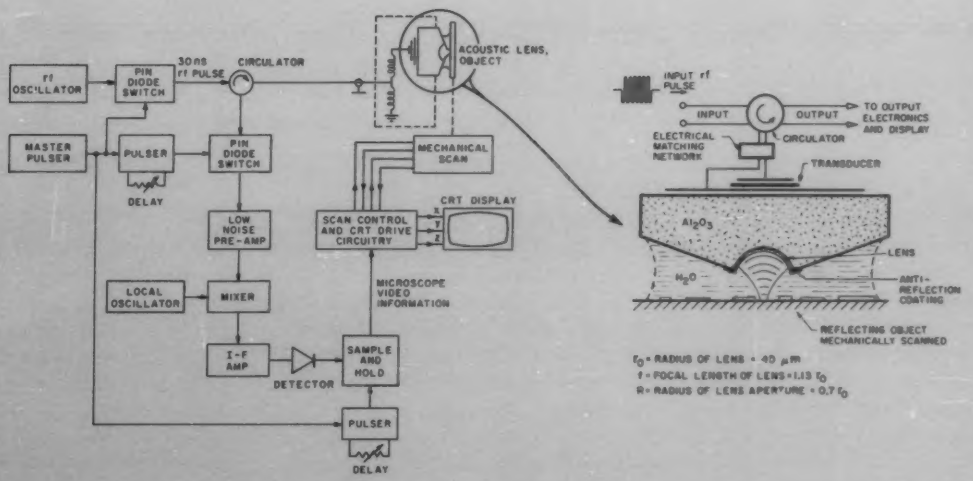
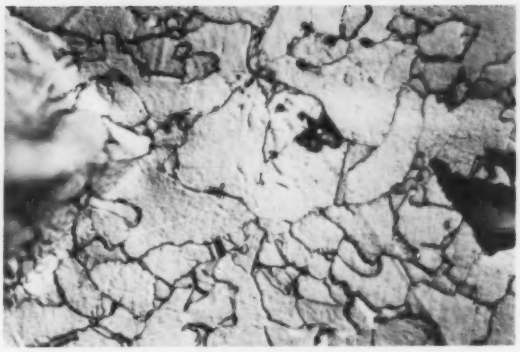


Figure 3: *The schematic of the overall electronic system (3a) used in the scanning microscope and the detail of the principle elements of the acoustic lens configuration (3b).*

wavelength in diameter at the focal plane. The energy outside of this region is negligible. There is no "field of view", which means we "see" one point at a time. Therefore it is necessary to build up the image point-by-point as it is done on a television or radar screen. Electrical fields can't be used to deflect acoustic beams in the manner of an electron beam in a CRT. Therefore this scanning is carried out with mechanical means since there is no other way to do it. This scanning slows the system to the point where several seconds are required to form the image.

What is the ultimate in resolution? This is determined by the diameter of the beam at the focal plane. In turn, the beam diameter is controlled by the wavelength of sound. Thus we plan to simply move to higher frequencies where the wavelength is shorter, but there are problems that have to do with the liquid in the acoustic cell which must be overcome along the way. Even the best of liquids are highly absorbing. In water heated to 60°C, 90 percent of the energy at 2 GHz is absorbed in a distance of 25 microns. This is one-third the diameter of a human hair. If we are to operate at that frequency the path length in the liquid cannot be much greater than this. Since the path length is about equal to the lens radius, a great deal of time is devoted to making small lenses. The current instruments operating at 2 GHz have lenses with radii in the range of 25 - 35 microns and the wavelength is 0.75 microns. The acoustic absorption increases as the square of the frequency, so to improve the resolution by a factor of two by moving from 2 GHz to 4 GHz it would be necessary to decrease the path length (i.e., the lens radius) by a factor of four. The lenses that we now have are small, and a reduction by a factor of four will be difficult. Therefore we look to other liquids.

The acoustic microscope under investigation with the ONR program is based on cryogenic fluids such as nitrogen, argon, and helium. The first two (nitrogen and argon) are interesting since the wave velocity is about one-half that of water. Some increase in resolution is possible but the absorption is still high. It is liquid helium that holds promise of dramatic improvements. The velocity of sound is six times smaller than it is in water and the sound absorption is strikingly different. When helium is cooled below 1°K, the absorption decreases as the fourth power of the temperature. At 0.1°K it is reduced by a factor of 10,000. The liquid is virtually transparent to sound at these temperatures, and the limitation literally disappears, which should allow us to operate with frequencies much greater than those used in the water instrument. This combination of low velocity which gives us a shorter wavelength for a given frequency and low absorption which permits us to work at higher and higher frequencies should one day lead to the construction of an acoustic instrument with a resolving power that is competitive with the scanning electron microscope.



(a) Optical

(b) Acoustic

Figure 4: *The surface of a low alloy steel after it has been polished and chemically etched. The fields are for different portions of the same specimen. The grain boundaries are clear in the optical image (4a), but the reflection contrast is identical for each grain. In the acoustic micrograph (4b), different grains appear with different contrast and we attribute this to various orientations of each grain.*

Helium is an important fluid on which future naval systems will depend. Two such systems are based on the superconducting properties of certain metals when they are immersed in liquid helium. Motors with cooled superconducting windings are being investigated in research laboratories and it is likely that they will one day appear in our ships. The cryogenic computer is also based on superconducting devices. It holds promise of increasing the speed of computing in advanced systems.

It is important to learn about the physical properties of these cryogenic fluids, which will be done with

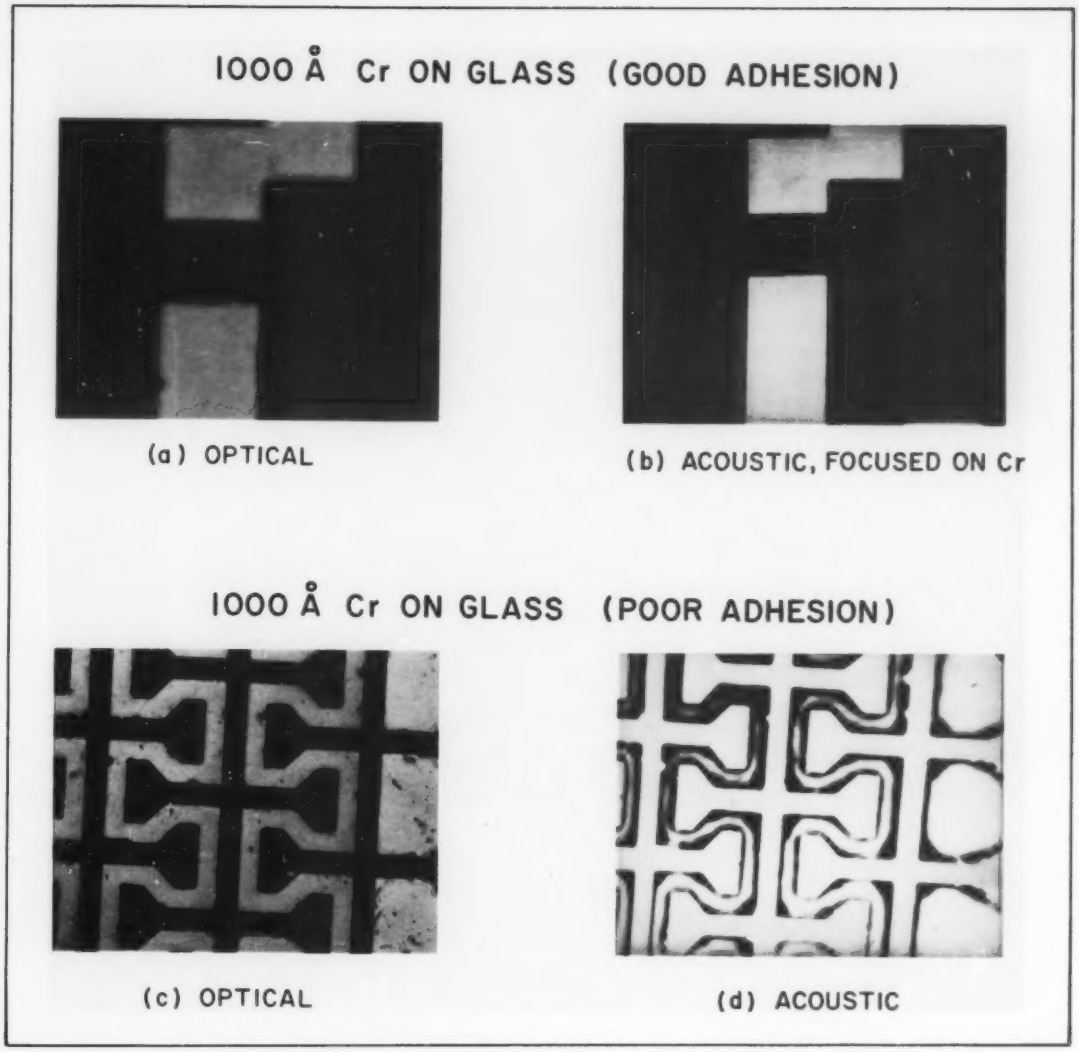
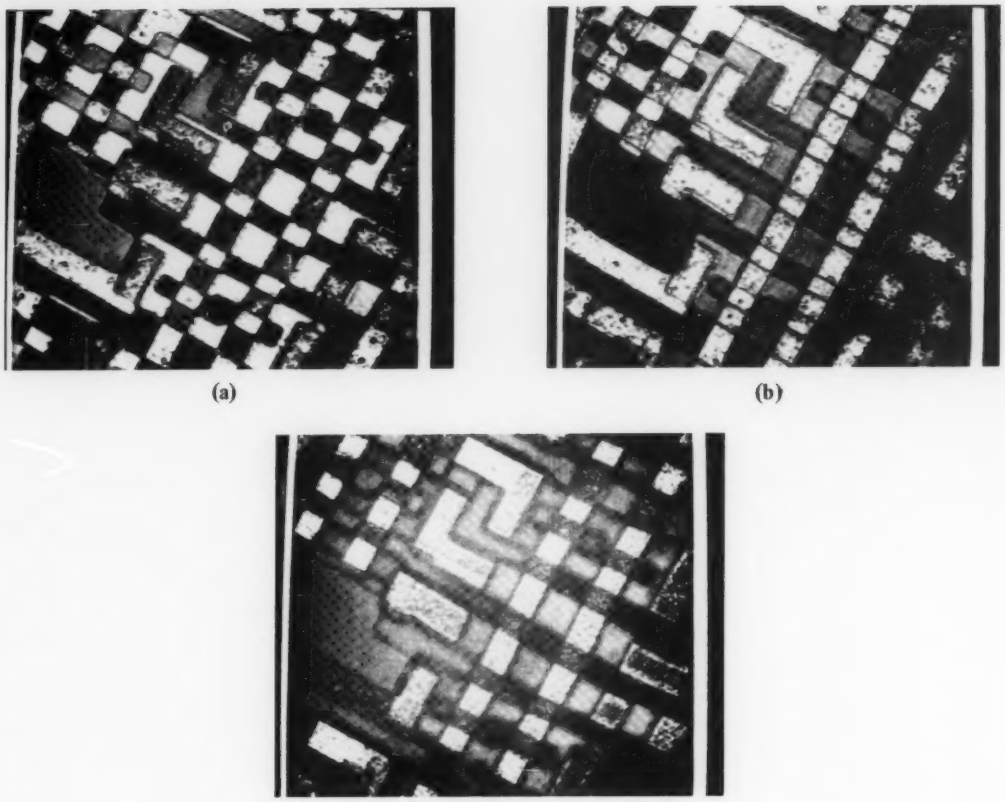


Figure 5: *A comparison of the adhesive properties of chromium on glass. In (a) and (b) we see the uniform reflectivity of the chrome pads in both the optical and the acoustic micrographs. In (c) and (d) we see the comparison for a poorly adhering film. The optical image is still uniform but the acoustic image is inhomogeneous–with the regions of poor adhesion appearing dark.*

probes of high-frequency sound. This research is primarily directed toward an imaging instrument with a resolution that exceeds that of the optical microscope. Such an instrument is needed because the technology of integrated circuits is advancing toward smaller and smaller circuits–some are below the limit of what we can see optically. It is important to find new instruments for imaging and for studying these fabricated circuits. The electron microscope (SEM) can do this, but it is limited to the uppermost layer whereas the acoustic microscope can probe some of the "in-depth" structure that lies beneath the top surface. In metals, an ever-increasing resolution will permit the investigators to study finer and finer detail. This, too, is imperative as we develop modern, high-strength materials. Polymers are used extensively for multiple purposes. They, too, are studied with microscopes, and their nature is such that there are numerous areas open for investigation with our acoustic probe–areas not suited for other microscropes.



(a) (b)

(c)

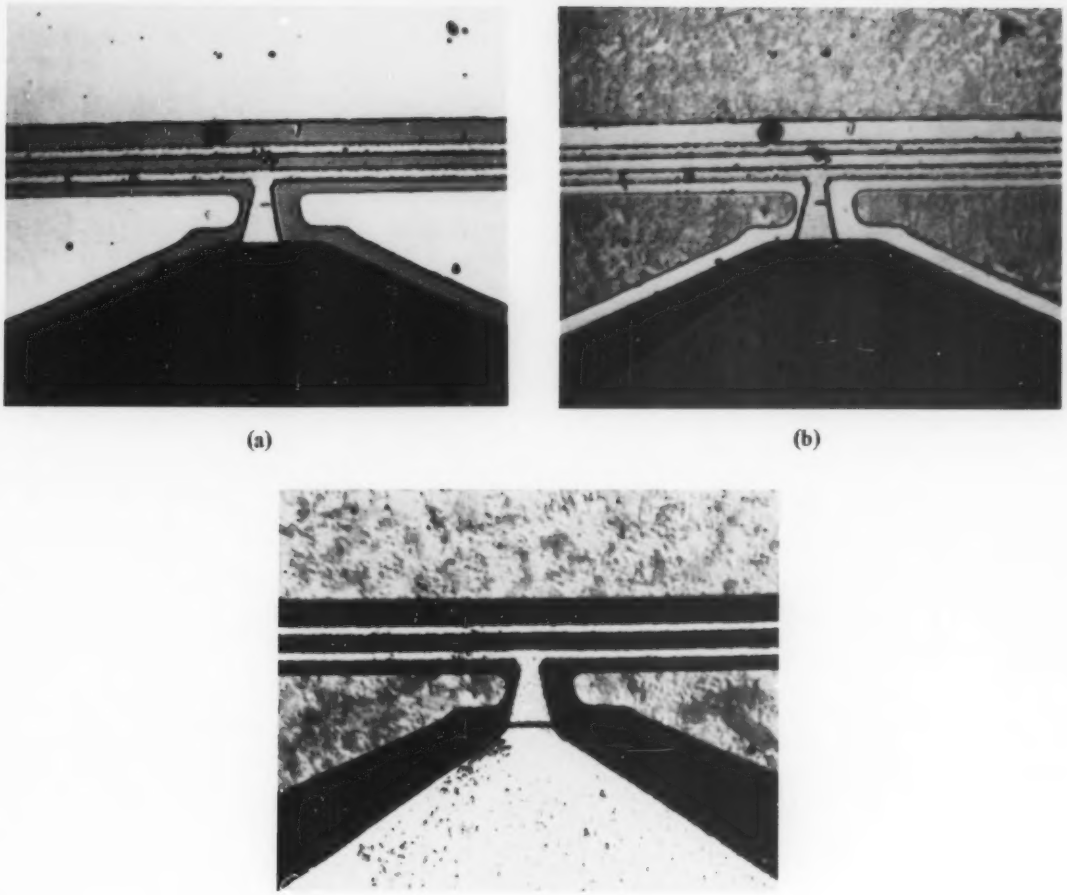
Figure 6: *The acoustic appearance of integrated circuits in the acoustic microscope. For a given element the acoustic contrast changes and even reverses as the sample is moved to different focal positions. The changing contrast is a function of the layering structure of these devices and a careful comparison will reveal information on the elastic properties of the layers.*

The Present State of the Art

The program on cryogenic microscopes is relatively new and the most interesting results still lie ahead. But there is enough information in the micrographs recorded with the room temperature instrument to illustrate the power and versatility of acoustic microscopy. Some examples that we have chosen for this section will demonstrate the present state of this field.

We have argued that acoustic microscopes can be used to monitor small changes in the elastic properties of the surface under observation. The first example, as in Figure 4, is a sample of low-alloy steel. We show the comparative images as found in both the optical and acoustic microscopes. The surface has been etched so as to reveal the boundaries of the individual grains. The regions between the grains etch at a different rate and we therefore see, in the optical image, the topography or contour of the surface. The optical reflectivity of each grain is quite similar. In the acoustic micrograph the reflectivity of each grain shows a wide variation as revealed in the change of contrast. We attribute this to a change in the orientation of the crystallographic axis of the grains. This property is termed "texture." It is important in determining the deformability of sheet material.

Figure 5 illustrates a case where the spatial variation of surface elasticity is important, deposits of chromium



(a) (b)

(c)

Figure 7: The acoustic micrograph from the cryogenic microscope. The liquid used here is argon near 80° K. The sample is a field effect transistor built on a wafer of gallium arsenide. The width of the two parallel lines is one micron. The two acoustic images (a) and (b) are taken at different focal positions. The optical comparison is given in (c).

layers on glass. The metal films are 1500 Å in thickness and are commonly used as masks in the photolithographic processes used with integrated circuits. The masks are used repeatedly and it is important to know the adherence of these films to the glass substrate. In Figure 5(a), we see a film that has good adherence, and the reflectivity of the film is uniform in both the optical and acoustic images. Contrary to this, Figure 5(b) shows that poorly adhering films have a different appearance in the acoustic micrographs. The edges of the metal films are dark and quite different from the central region. We believe that in these samples the edges of the film are not firmly attached to the glass substrate. This lack of adhesion is nowhere evident in the optical images. The adhesion of films is a problem that is encountered in many places. This property of acoustic micrographs may be a unique tool for determining this parameter.

The acoustic images of integrated circuits are shown in Figure 6. The differing contrasts appear as the position of the surface is moved from the focal plane, and the reversal of contrast for a given region is due to the layering that is part of these structures. If in some way the

layering was altered as in a defective device, the contrast would change and the defect would be revealed.

The micrograph of Figure 7 illustrates two points: firstly it demonstrates resolving power, and secondly it demonstrates that cryogenic fluids are viable liquids for acoustic microscopes. The device is a field-effect transistor built on a gallium arsenide wafer. The two parallel lines are gate electrodes which control the current flow. They are one micron in width and spaced by one micron. We can agree that these lines are clearly resolved. The wavelength of sound used in recording this image was 0.6 micrometers. It is therefore reasonable to resolve one-micron lines.

The instrument used here was immersed in liquid argon near 80°K. Since this is the first cryogenic fluid that we have used, the results are gratifying. We believe that it is but the first step in our program for high-resolution microscopy in liquid helium. A helium instrument cooled to 2°K was used to produce images with a high degree of detail, but the loss in helium was large at 2°K, larger than in water. It will be necessary to cool to 0.2°K before the full potential of fluid helium is realized.

In Summary

Acoustic microscopy is now established as a form of imaging that furnishes information in the microscopic world that complements and adds to the material gathered with the optical and electron microscopes. The instrument is now being investigated in several laboratories distributed throughout eight countries. It is being considered for commercial development by one major instrument firm.

The resolving power is comparable to that found in optical micrographs and this is expected to improve as we move forward with the cryogenic instrument. The instrument, as it now stands, is well suited for investigating the details of integrated circuits and other microelectronic devices such as found in integrated optical systems and modern memory devices. It has proven to be effective in the microscopic study of a variety of materials ranging from steel to plastics to organic rocks, such as coal and shale.

Conventional forms of microscopy include both the optical and the electron microscopes. In each of these instruments advances continue to be made with new modes of operation and new methods for control and display of the images. Their use is increasing on an ever-widening base since the improvements permit one to tackle more and more difficult problems. It is our hope that the acoustic microscope will find its place as a third form of microscopy; one that can be used to collect information that complements that gathered by the other two. When this hope is realized, it will mean that X-rays remain as the next frontier. One day that form of radiation also will be incorporated into a focused imaging system. ■

References

1. C. F. Quate, "The Acoustic Microscope," *Scientific American*, 241, pp. 62-81 (October 1979).

2. L. W. Kessler and D. E. Yuhas, "Acoustic Microscopy – 1979," *Proc. IEEE*, 67, pp. 526-535 (April 1979).

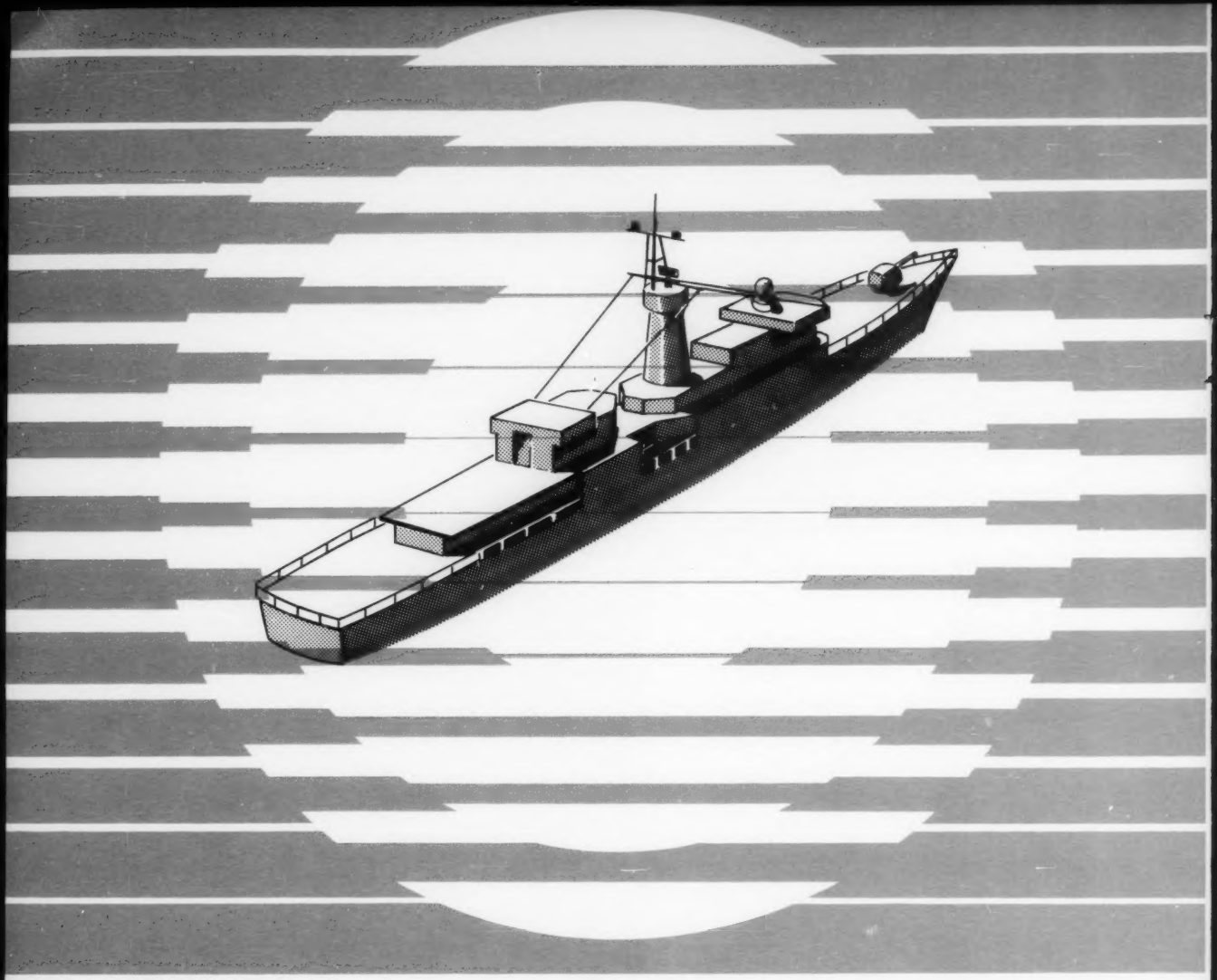
3. C. F. Quate, A. Atalar, and H. K. Wickramasinghe, "Acoustic Microscopy with Mechanical Scanning – A Review," *Proc. IEEE*, 67, pp. 1092-1114 (August 1979).

4. J. Heiserman, D. Rugar, and C. F. Quate, "Cryogenic Acoustic Microscopy," *J. Acous. Soc. Am.*, 67, pp. 1629-1637 (May 1980).

profiles in science

Dr. Alan J. Heeger is Professor of Physics and Director of the Laboratory for Research on the Structure of Matter at the University of Pennsylvania. For many years Professor Heeger has been a contractor for the Office of Naval Research which helped fund his pioneering research in the area of highly conducting organic solids. His current research interests lie in the development of the fundamental physics which determine the electronic and magnetic properties of organic and organometallic solids, with the eventual goal of making these novel materials available for technological applications.

Professor Heeger, who has published 200 articles in scientific journals, was the Morris Loeb Visiting Lecturer in Physics at Harvard in 1973 and was a National Academy of Sciences exchange scholar to the U.S.S.R. in 1977.



Frequency Independent Beamforming *

*ACKNOWLEDGEMENT

This article is based on a presentation made at the Naval Underwater Systems Center in Newport, Rhode Island, to Dr. James H. Probus, Director of Naval Laboratories, Dr. Tibor G. Horwath, Director of Naval Technology and Mr. Carey D. Smith of the Naval Sea Systems Command on April 19, 1979 and describes studies carried out on Project FLYEYE (a multi-objective sonar technology development program) under joint sponsorship by the Office of Naval Research, the Naval Underwater Systems Center and the Naval Sea Systems Command.

by

Robert L. Sternberg
Office of Naval Research,
Eastern/Central Regional Office
Boston, Massachusetts

and

Warren A. Anderson
Naval Underwater Systems Center
New London, Connecticut

Introduction

In a number of critical areas of radar and sonar technology and also in spread spectrum communications and perhaps even in directed-energy micro-wave weapons applications, many advantages would attend the development of frequency-independent, multiple or steerable beamforming directional antennas. Such antennas provide a means of producing, on a multiple or steerable beam basis—but with a single radar or sonar antenna—transmitting and/or receiving beams whose beamwidths, directivity indices, antenna gains and beam shapes remain invariant as the signal frequency varies over a decade of more of the frequency spectrum.

Early interest in such antennas with somewhat more limited objectives developed in the electromagnetic communications realm for signal jamming and anti-jamming purposes and related electronic warfare applications following Dyson's fundamental discoveries in the late 1950's (1,2). The first interest in such devices for sonar appears to have been motivated by commerical fishing applications and the very broad spectrum of sound radiated by some marine animals, as described for example by Tucker and Gazey (3) and shown in Figure 1. Thus frequency agility and signal security requirements drove the electromagnetic work, while the possibility of spectral identification of broadband targets observed passively with frequency-independent directional hydrophones, if realizable, gave promise in sonar of a new, more effective method of locating and identifying schools of the fish. In the underwater acoustics applications the important point was the elimination of the spectral distortion introduced by an ordinary directional antenna or hydrophone as compared to a frequency-independent, or constant beamwidth, antenna as described below.

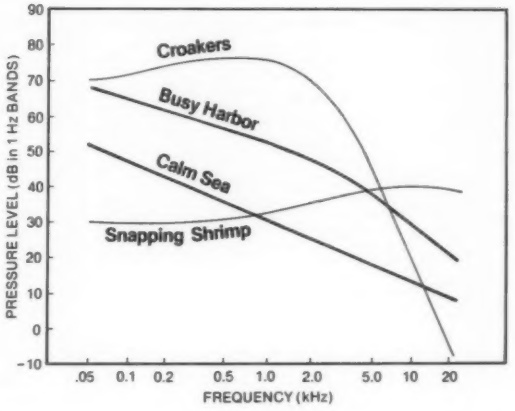


Figure 1. Spectral distribution of various background and fish noises in the sea. After D. G. Tucker and B. K. Gazey Applied Underwater Acoustics, *Pergamon Press 1966; borrowed with permission.*

Log-periodic and related antennas in the electromagnetic world have been able to form frequency-independent, constant beamwidth, beams on a single non-steerable beam basis for many years as described, for example, by Rumsey (4), and several theoretical and experimental papers related to frequency-independent, and in some cases steerable, beamforming in underwater acoustics based on a variety of antenna array or analog techniques have appeared in the literature. Listed in chronological order and especially noteworthy among the latter are the works of Morris and Hands (5), Morris (6), Hixson and Au (7), Sternberg, Anderson and Stevens (8), Rogers and Van Buren (9) and Jarzynski and Trott (10) which are briefly reviewed in the present study.

***Dr. Robert L. Sternberg** is a Mathematician and Physical Science Administrator at the Office of Naval Research, Eastern/Central Regional Office in Boston, Massachusetts. Dr. Sternberg, who is credited with many publications and several patents, has spent much of his career in radar antenna research as well as applied mathematics and engineering in various defense industries and research establishments.*

Dr. Sternberg's current research in nonspherical geometrical optics and frequency independent acoustic antennas have involved him with Project FLYEYE at the naval Underwater Systems Center as part of his formally assigned Navy duties since 1975.

***Warren A. Anderson** is an Electronic Engineer at the Naval Underwater Systems Center, Newport, Rhode Island. Mr. Anderson's career in electronics began in 1939 with RCA Communications—Terminal Facilities Laboratory.*

During his 40 years of service to the electronic arts, Mr. Anderson has been awarded more than 27 patents. He was elected a Senior Member of the Institute of Electrical and Electronic Engineers in 1963. Mr. Anderson has been associated with Project FLYEYE since 1976.

Of the several different sonar-related approaches to frequency-independent or constant beamwidth beamforming noted, all except the work of Sternberg, Anderson and Stevens were based on, or derived in one form or another from fundamental planar or spherical array antenna concepts, with the beamforming done electrically by signal processing. Using such antenna array techniques, some promising but rather limited experimental results in frequency-independent beamforming were reported in the paper by Morris, Hands, Hixson and Au but theoretically computed results only were given by Rogers, Van Buren, Jarzynski and Trott.

Using a novel, more or less analog technique in an empirical approach, Sternberg, Anderson and Stevens (8) demonstrated experimentally a quite different method of frequency-independent beamforming on a steerable beam basis using a low impedance Sylgard rubber acoustic lens of non-spherical design together with a new, high impedance, stainless steel device—called an acoustic filter plate—to form passive receiving beams directly without the use of array-processing techniques. More particularly, with the acoustic lens and filter plate components mounted in the configurations illustrated in Figure 2 with a movable hydrophone back of the acoustic lens, several preliminary frequency-independent scanning beam patterns were obtained in the latter part of 1976, and the very much better beam patterns shown in Figure 3 were obtained in experiments conducted early in 1978. The antennas in each case were used only as receivers.

In the next section we present some of the special advantages and properties of frequency- independent or constant beamwidth beamforming for various, mainly sonar-related, applications. In the third section of the paper, we review in chronological order some of the early sonar array work of Morris, Hands, Hixson and Au, noted above. The fourth and fifth sections present an update of the acoustic lens-acoustic filter plate work of Sternberg, Anderson and Stevens and summarize the status of the later array work of Rogers, Van Buren, Jarzynski and Trott. The sixth section presents a new, not previously published, "mosaic lens array antenna" concept which the present authors believe offers some new and potentially useful and interesting alternatives as an approximately freuqency- independent, essentially constant beamwidth, multiple or steerable beam antenna for either transmission or reception and ap-

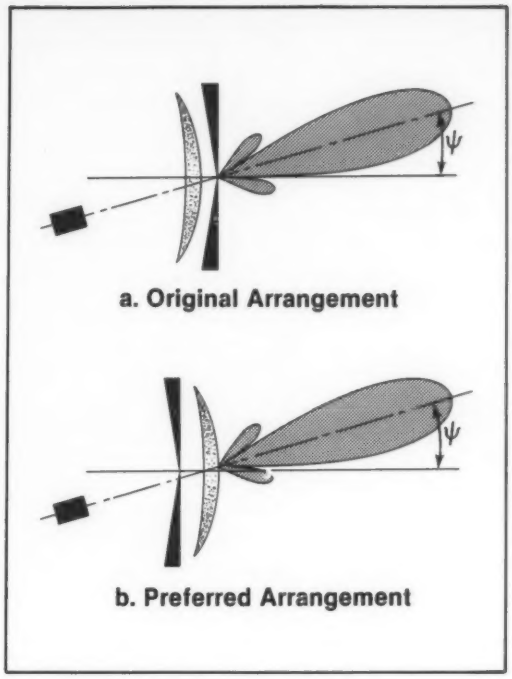


Figure 2. *Alternative acoustic lens-acoustic filter plate mounting arrangements; (a) original arrangement and (b) preferred arrangement.*

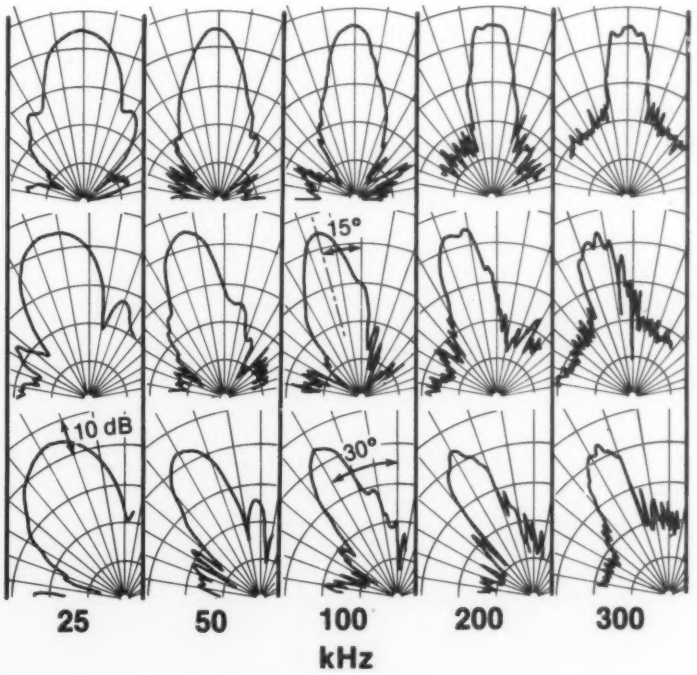
plicable either to radar or sonar. In the last section we conclude with some observations of directions in which we believe research and development in this field will move in future years.

Advantages of Frequency Independent Beamforming.

The special advantages and attributes enjoyed by frequency-independent, constant beamwidth, directional radar or sonar antennas, whether of a multiple or steerable beam type or not, can all be summarized by noting that such antennas eliminate the angular dependence of the spectral distortion introduced by an ordinary frequency-dependent, variable beamwidth, directional antenna when the transmitted or received signal contains more than a single frequency term. However, how this comes about can be described in various ways.

As a first illustration consider a multiple beamforming, freqeuncy-dependent, variable beamwidth radar or sonar antenna such as an ordinary Luneberg

Figure 3. *Experimentally measured frequency-independent acoustic lens-acoustic filter plate receiving beam patterns taken at various frequencies and scan angles showing approximately constant beamwidths of about 12° at frequencies above 25 kilohertz.*



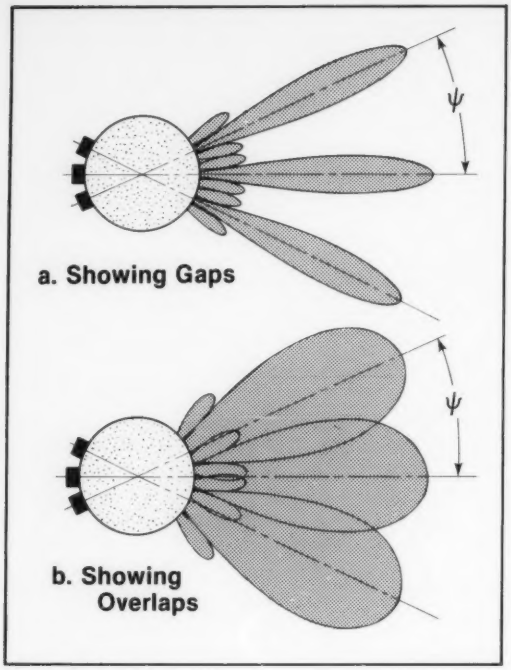


Figure 4. *Multiple beam formation with an ordinary Luneberg lens equipped with a retina of wideband transmitting or receiving elements; (a) showing gaps between the beams at high frequencies and (b) showing overlaps of the beams at low frequencies.*

lens equipped as in Figure 4, with a retina of wideband transmitting or receiving elements rigidly mounted on its focal surface as shown. Such an antenna produces a multiplicity of transmitting or receiving beams in a number of fixed directions relative to the antenna axis. As the frequency changes both the beam directions and the aperture openings—which here are the same as the lens diameter D—remain unchanged. Since the beamwidths are all the same constant multiple of the ratio $R = \lambda/D$, and in fact are given approximately by the formula

$$B.W. \cong 65\ \lambda/D$$

in degrees, it can be seen at once that the beams formed by the antennas are narrow at high frequencies and broad at low frequencies. Consequently as the frequency varies, gaps develop betwen the beams when the frequency is high and the beams overlap when the frequency is low as illustrated.

With a frequency-independent, constant beamwidth antenna of course no such gaps or overlaps occur as the frequency changes.

Consider next the effects on the acoustic spectrum of a passively observed sonar target in ASW. Suppose the target is radiating either signals or noise over a wide band of frequencies while it is observed in the clear by a scanning, sonar-receiving beam formed by a frequency-dependent, variable beamwidth antenna. As the beam scans by the ASW target, as illustrated in Figure 5(a) the amplitudes of the various frequencies in the spectrum radiated by the target are continuously distorted or modulated at the receiver by the constantly changing beamwidth and beam shape of the antenna, as a function of spectral frequency. The result is that the amplitudes of the high-frequency terms are heavily attenuated by the sonar antenna except when the target is on, or nearly on, boresight, while low frequencies are picked up even when the target is far off the beam axis. In fact the true spectral content of the target signals arriving in the vicinity of the antenna is received by the antenna without amplitude modification as a function of frequency only when the beam from the antenna is pointed directly at the target, i.e., only at the transitory instant when the angle-off-boresight, $\triangle\Psi$ in the illustration, is zero.

With a frequency-independent, constant beamwidth, sonar antenna on the other hand no such spectral distortion occurs, because, as shown in Figure 5(b), all frequencies arriving from the target are received with essentially the same antenna gain, independent of the frequency throughout the dwell time of the constant beamwidth beam on the target, regardless of whether the beam is perfectly aligned on boresight with the target or not.

As a final illustration consider the situation that occurs when an extended target such as the sea floor or a large school of fish is actively observed with a frequency-dependent, variable beamwidth downward-looking sonar with the same transducer serving as both the transmitter and the receiver. In this case, as shown in figure 6(a), the size of the target patch ensonified by the sonar antenna varies inversely with the signal frequency. This means that only a very small patch on the extended target is illuminated at the highest frequency transmitted, while ever-larger patches are irradiated with sound at lower and lower frequencies. Thus the spectral content of the signal reflected by the target and received then is not characteristic of any single part of the extended target, but consists instead of high frequency terms reflected by the small central patch, intermediate frequency terms reflected by somewhat larger circular patches concentric with the first, and low frequency components reflected by still larger circular patches concentric with the first two. For things like ocean resources prospecting or buried mine hunting in the

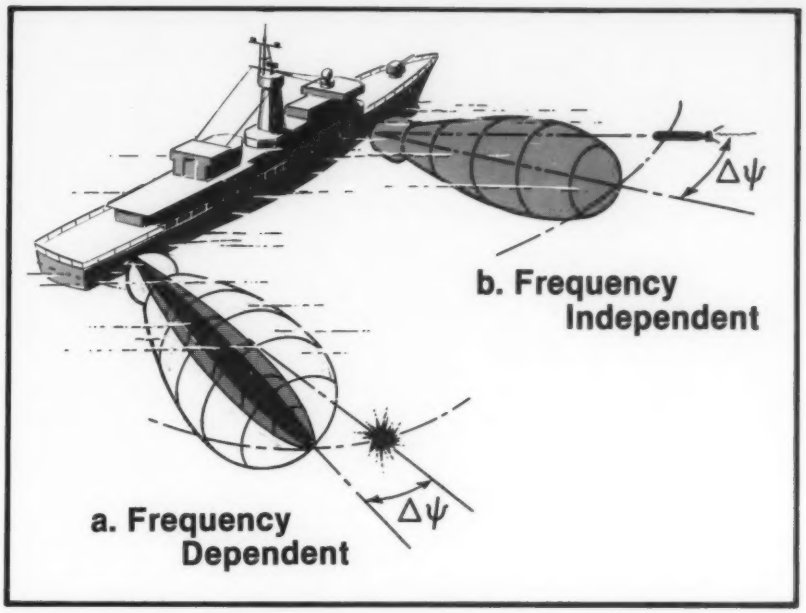


Figure 5. ***Passive surveillance of a wideband noisy point source ASW target in the clear; (a) with a frequency-dependent sonar antenna and (b) with a frequency-independent constant beamwidth sonar antenna.***

sea floor with a downward-looking sonar as in the illustration, spectral characterization of individual parts of the extended sea floor target by their frequency- dependent reflectivity properties is then made highly complicated, if not totally impractical, by the "smearing" of the reflected signal spectrum due to the spacial extension of the target combined with the automatic broadening of the sonar beam at low frequencies.

As in the previous illustration, the situation is much improved by using a frequency-independent, constant beamwidth sonar antenna as shown in Figure 6(b). A given patch on the extended target is now equally ensonified by all frequencies. Since the same patch also provides the reflected signal, it follows that the character of the bottom material making up the patch can be correlated with the returned spectrum for spectral identification purposes, provided that sufficiently extensive archival data have previously been collected. For acoustic studies of the ocean floor or for ocean resources research this is a most significant property.

Early Sonar Array Methods

Perhaps the first substantial study of frequency-independent or constant beamwidth beamforming for sonar applications to appear in the literature was the work of Morris and Hands (5). After first outlining some elementary ways in which frequency- independent beams might be formed by array processing techniques using shading filters, resonant arrays or time-delay methods, Morris and Hands analyzed and developed a method proposed earlier by Tucker (11), based on synthesis of directional beams from elementary (sin x)/x diffraction patterns using a multitapped delay line. With this approach, if ϕ denotes the total phase shift of the delay line and as usual $x = \sin \theta$ then the nominal directional diffraction pattern takes the form

$$D = \frac{\sin (x \pm \phi/2)}{x \pm \phi/2}$$

according to their conclusions. The beam produced is almost, but not quite, constant in beamwidth. More precisely, theoretically computed beam patterns were obtained predicting constancy of beamwidth over a nine-to-one frequency change ratio, with less than ± 10 percent change in beamwidth. However, owing to a number of practical difficulties encountered,

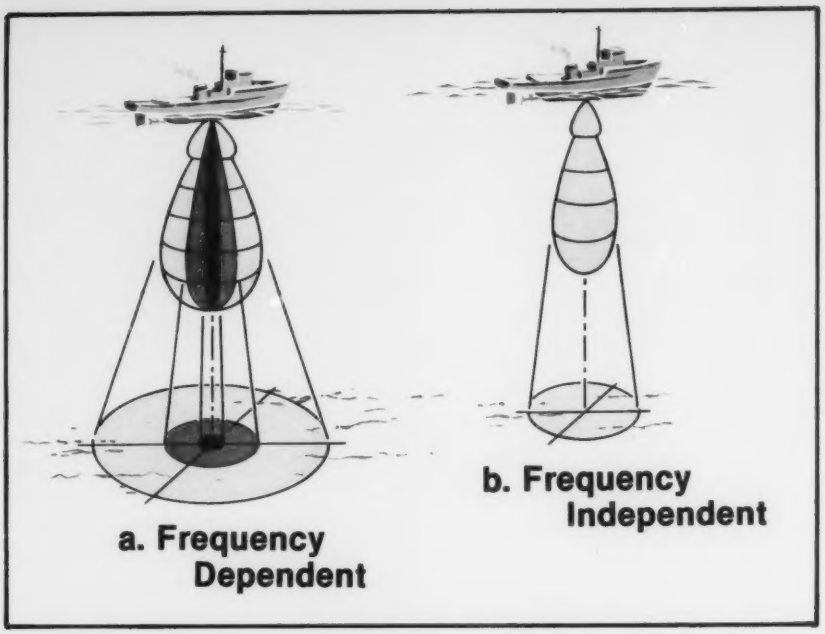


Figure 6. *Active surveillance of an extended sea floor target area as in ocean resources prospecting or buried mine hunting; (a) with a frequency-dependent sonar antenna and (b) with a frequency-independent constant beamwidth sonar antenna.*

Morris and Hands were able to verify these results experimentally only over a very much smaller range of frequencies, covering in fact less than a two-to-one excursion in that parameter.

The next advance, again carried out mainly on a theoretical basis, was the investigation by Morris (6) of various curved arrays. While reporting some promising but not extensive experimental progress involving curved line arrays, Morris went on to an interesting analysis of the concept of a "twisted" planar, or hyperbolic paraboloidal, array. Here the array elements are mounted on a twisted plane which approximates a hyperbolic paraboloid in the neighborhood of its saddle point. The mechanism by which such an array forms a frequency-independent beam is too complicated to describe here but was explained in some detail by Morris. He computed three-dimensional plots of the main lobes of the diffraction patterns formed by such a twisted array built in the form of a square, with the length of side $n\lambda$, for $n =$ 2, 3, 5, 8, 10, 20 and 40. However full experimental confirmation of the apparently excellent frequency-independent,beamforming characteristics of this type of antenna appears not yet to have been attempted either by Morris or by others.

About ten years after the early researches of Morris and Hands (5) on frequency-independent or constant beamwidth antennas, Hixson and Au (7) carried out a theoretical and experimental investigation of a "nested" line array of receiver elements—in their case microphones operating in air—arranged somewhat as shown in Figure 7. Here the array element spacing is such that each successive element, counted from the ends of the array towards the array center, is spaced half as far from the center as the preceding element. For the array consisting of the seven microphone elements E'_3, E'_2, E'_1, E_0, E_1, E_2, E_3 shown in the illustration, with each microphone in the array responsive to the frequencies indicated, it follows that if the outputs of the array elements are suitably filtered and processed, then the two sub- arrays consisting respectively of the two sets of elements

$$E'_3, E'_2, E_0, E_2, E_3$$

and

$$E'_2, E'_1, E_0, E_1, E_2$$

can be made to form identical receiving beams at the two frequencies f_0 and $2f_0$, since each sub-array is an image of the other scaled by the corresponding

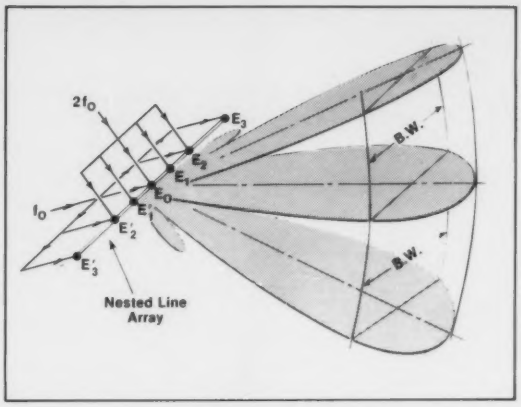


Figure 7. *Nested line array concept of Hixson and Au for formation of a frequency-independent fan beam with the same azimuth beamwidths at frequencies f_o and $2f_o$ and approximately constant azimuth beamwidths at intermediate frequencies.*

wavelength. By increasing the number of elements in the array center and the number of frequencies to which each element responds while keeping the total length of the array constant, the same beams may be formed not merely at the frequencies f_o and $2f_o$ but, more generally, at all of the frequencies of the form

$$f_o, 2f_o, 4f_o, 8f_o, 16f_o$$

and so forth. The main limitations here are those caused by the bandwidths of the individual array elements and the extent to which the necessary additional elements can be "packed" into the array center to form new sub-arrays each of which is a scaled image of the others. Moreover, by additional filtering and smoothing processes the overall bandwidth of the array can be made continuous, with beamwidths varying somewhat periodically as the frequency changes, but with the original beam shape and beamwidth being recaptured each time the frequency is doubled. Finally, it is clear that the same concepts can be extended to two-dimensional arrays without inherent difficulty. In addition to their fundamental theoretical investigations Hixson and Au also carried out a limited, but nevertheless essentially successful, program of experiments verifying their basic concepts.

Although substantial progress was reported by these writers on fundamental antenna array techniques for frequency-independent beamforming in acoustics there was no real scrutiny in these early programs of the possibility of extending the methods studied to the problem of multiple or steerable beamforming with constant beamwidths although it was noted, and is quite apparent, that some of the methods could readily be extended in that way.

Lens—Filter Plate Method

The next step taken toward development of frequency-independent or constant beamwidth antennas for sonar applications, this time carried out on a steerable receiving-beam basis, was the work of Sternberg, Anderson and Stevens (8) on the acoustic lens-acoustic filter plate method. In this approach to the problem utilizing one of the mounting configurations shown in Figure 2, with a movable hydrophone scanning the focal surface back of the lens as shown, steerable frequency-independent receiving beams are formed automatically in analog fashion, by the combined actions of the lens and the filter plate—the former providing the focussing and the latter acting as an aperture stop. More particularly, because of the geometrical shape of the filter plate—planar on one surface, conical on the other and infinitely thin at its center—and because of the high impedance properties of the stainless steel material of which it is made as described in the reference cited, the filter plate acts automatically to vary the effective aperture diameter D_E of the combined lens-filter plate device in proportion to the wavelength λ, thereby keeping both the ratio $R = \lambda/D_E$ and the beamwidth $B.W. \cong 65\lambda/D_E$ constant.

The first set of such steerable, frequency-independent receiving beam patterns, taken in 1976 with a seven-inch diameter Sylgard rubber lens of nonspherical design and a slightly larger diameter stainless steel filter plate mounted in the configuration shown in Figure 2(a), showed more or less constant beamwidths of about 12 degrees over a band from about 25 kilohertz to 250 kilohertz. Scanning performance was good out to about ± 15 degrees off the lens-filter plate axis but deteriorated rapidly at angles of about ± 30 degrees or more off-axis. Later beam patterns taken early in 1978 with the lens-filter plate combination arranged in the configuration shown in Figure 2(b)—in particular with the filter plate back of the lens instead of in front of it— showed much improved beam shapes that remained well formed right out to ± 30 degrees off-axis, as a comparison of Figure 3 with the previously reported patterns of Sternberg, Anderson and Stevens shows. Moreover, continued exploration of the lens-filter plate method, again using the mounting configuration of Figure 2(b), has shown that the method is suitable also for forming frequency-independent, constant beamwidth beams at reasonably high resolutions, as indicated by the beam patterns shown

in Figures 8 and 9, obtained late in 1979 and early in 1980, using filter plates of lesser angle-of-taper than before. In fact, these results strongly suggest that arbitrarily high resolution, or narrowness of the beams formed, can readily be achieved with filter plates of sufficiently low angle-of-taper, or peripheral thickness, as this opens up the effective aperture at each frequency to a larger fixed number of wavelengths than before. In this connection we may note that the effective aperture diameters D_E for the filter plates used in taking the patterns of Figures 3, 8 and 9 were about 5λ, 15λ and 30λ respectively.

sion, the lens-filter plate method should eventually be able to form not just one steerable frequency-indepedent receiving beam but should be able to form many such beams simultaneously so as to "cover" the observation cone continuously, without time sharing, with a multiplicity of fixed beams in different directions all having the same shape and beamwidth.

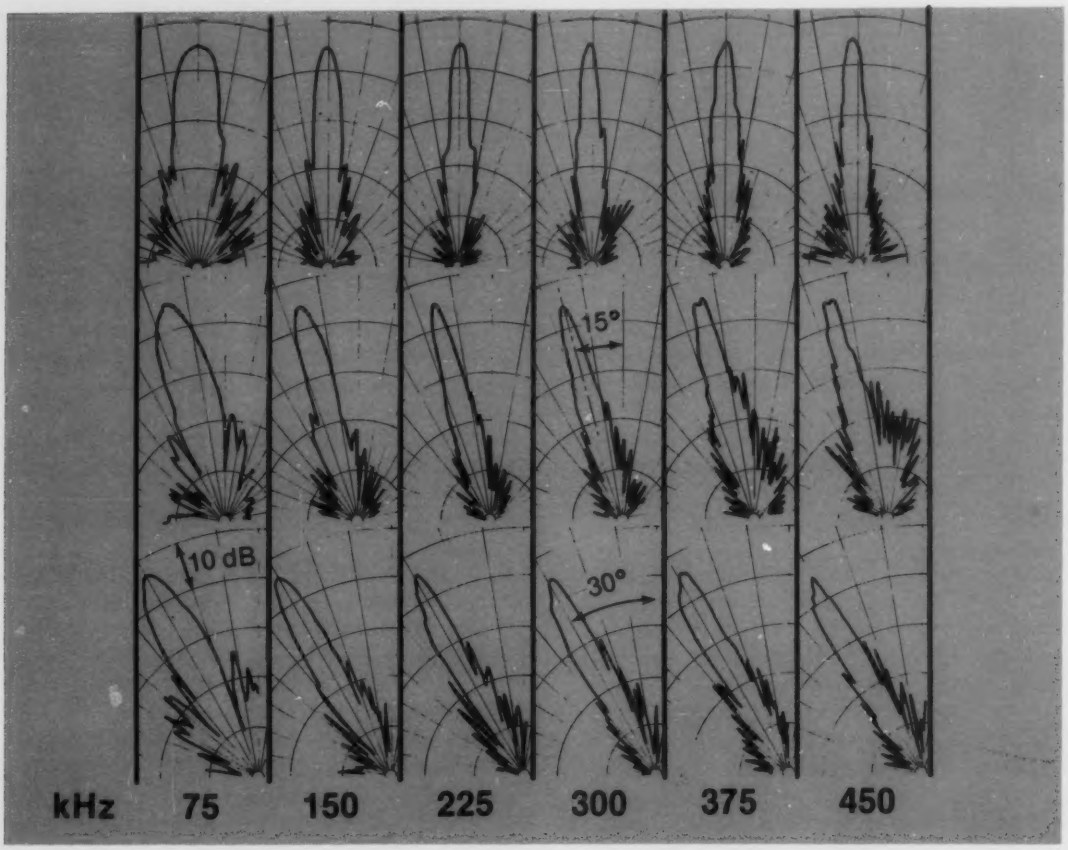


Figure 8. *Experimentally measured frequency-independent acoustic lens-acoustics filter plate receiving beam patterns taken at various frequencies and scan angles showing approximately constant beamwidths of about 4 at frequencies above 75 kilohertz.*

Finally, these results not only clearly demonstrate the capability of the lens-filter plate method to form steerable, frequency-independent, constant beamwidth receiving beams but also strongly suggest the eventual possibility of forming multiple beams of the same kind by replacement of the solitary moving hydrophones shown back of the lenses in Figures 2(a) and (b) with complete acoustic retinas of such hydrophones. With such an exten-

Spherical Array Methods

Appearing next, and closely related to one another, were the theoretical investigations of spherical array concepts for frequency-independent, constant beamwidth beamforming carried out by Rogers and Van Buren (9) and Jarzynski and Trott (10).

velocity on the surface of the transducer is of the form

$$u = u_0 P_\nu(\cos\theta)$$

where $P_\nu(x)$ is a suitably chosen Legendre function whose root of least angle occurs at $\theta = a$, then the antenna may be shown theoretically to enjoy uniform acoustic loading, to produce extremely low sidelobes,

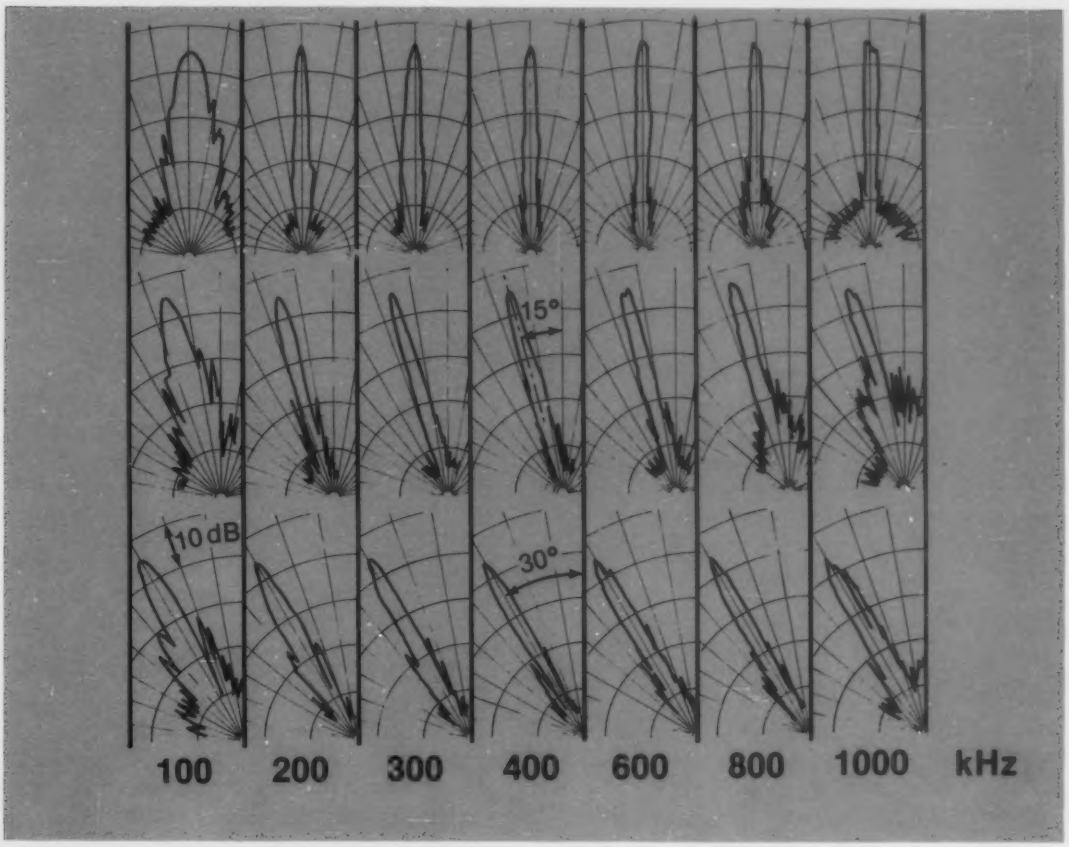


Figure 9. *Experimentally measured frequency-independent acoustic lens-acoustic filter plate receiving beam patterns taken at various frequencies and scan angles showing approximately constant beamwidths of about 2° at frequencies above 100 kilohertz.*

The array antenna concept studied by Rogers and Van Buren consists of an appropriately shaded spherical-cap transducer, or array of separate transducer elements closely approximating such a spherical-cap, as indicated in Figure 10(a). In this concept, if the angular extension of the spherical-cap transducer or array is $2a$ as illustrated, and the shading is such that the distribution of the normal

and to have an essentially constant beamwidth independent of the frequency whenever the latter exceeds a certain cutoff value. The transducer cap or array is mounted on a rigid acoustically-opaque sphere and is apparently intended to be used primarily as a transmitter although the same principles would clearly permit its use as a receiver.

In the related concept envisioned by Jarzynski and Trott the antenna is made up of an array of isophase, omnidirectional transducer elements mounted on a spherical surface much as before except that here—in principle although not in practical approximations—the array should completely fill out the sphere as in Figure 10(b). With shading again ex-

in Figures 8 and 9, obtained late in 1979 and early in 1980, using filter plates of lesser angle-of-taper than before. In fact, these results strongly suggest that arbitrarily high resolution, or narrowness of the beams formed, can readily be achieved with filter plates of sufficiently low angle-of-taper, or peripheral thickness, as this opens up the effective aperture at each frequency to a larger fixed number of wavelengths than before. In this connection we may note that the effective aperture diameters D_E for the filter plates used in taking the patterns of Figures 3, 8 and 9 were about 5λ, 15λ and 30λ respectively.

Finally, these results not only clearly demonstrate the capability of the lens-filter plate method to form steerable, frequency-independent, constant beamwidth receiving beams but also strongly suggest the eventual possibility of forming multiple beams of the same kind by replacement of the solitary moving hydrophones shown back of the lenses in Figures 2(a) and (b) with complete acoustic retinas of such hydrophones. With such an extension, the lens-filter plate method should eventually be able to form not just one steerable frequency-indepedent receiving beam but should be able to form many such beams simultaneously so as to "cover" the observation cone continuously, without time sharing, with a multiplicity of fixed beams in different directions all having the same shape and beamwidth.

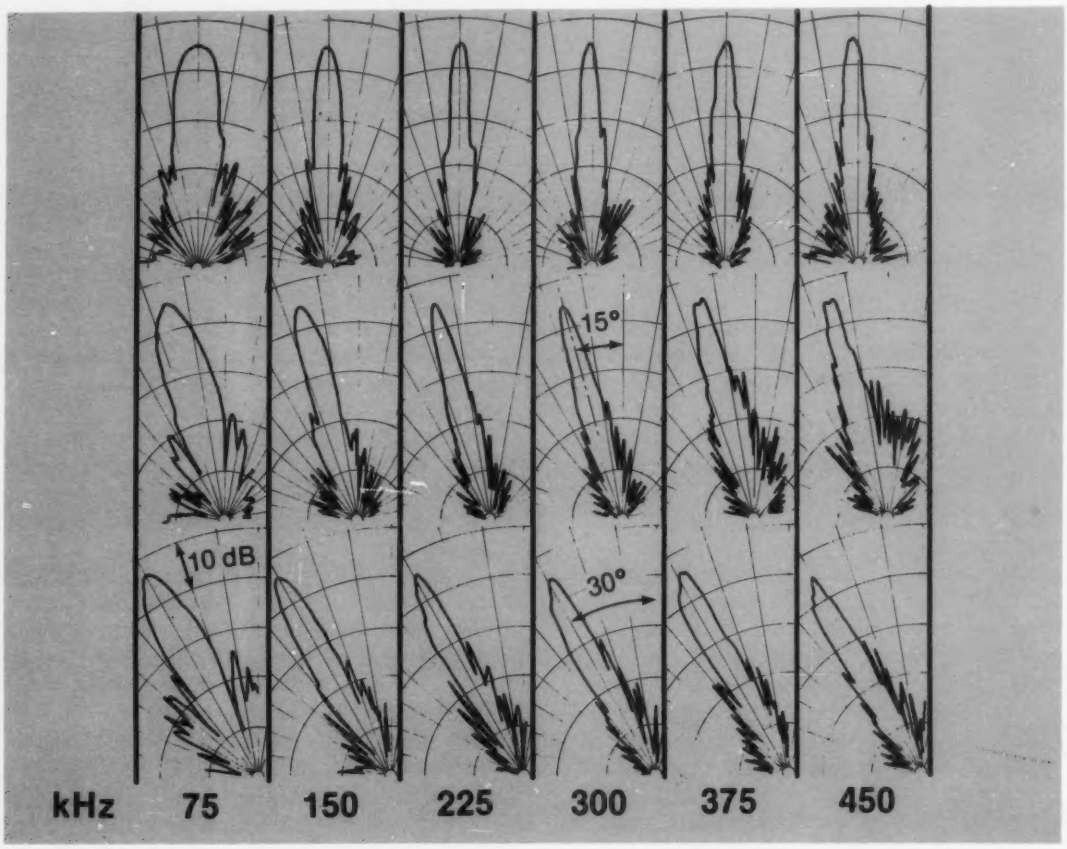


Figure 8. *Experimentally measured frequency-independent acoustic lens-acoustics filter plate receiving beam patterns taken at various frequencies and scan angles showing approximately constant beamwidths of about 4 at frequencies above 75 kilohertz.*

Spherical Array Methods

Appearing next, and closely related to one another, were the theoretical investigations of spherical array concepts for frequency-independent, constant beamwidth beamforming carried out by Rogers and Van Buren (9) and Jarzynski and Trott (10).

velocity on the surface of the transducer is of the form

$$u = u_0 P_\nu(\cos\theta)$$

where $P_\nu(x)$ is a suitably chosen Legendre function whose root of least angle occurs at $\theta = a$, then the antenna may be shown theoretically to enjoy uniform acoustic loading, to produce extremely low sidelobes,

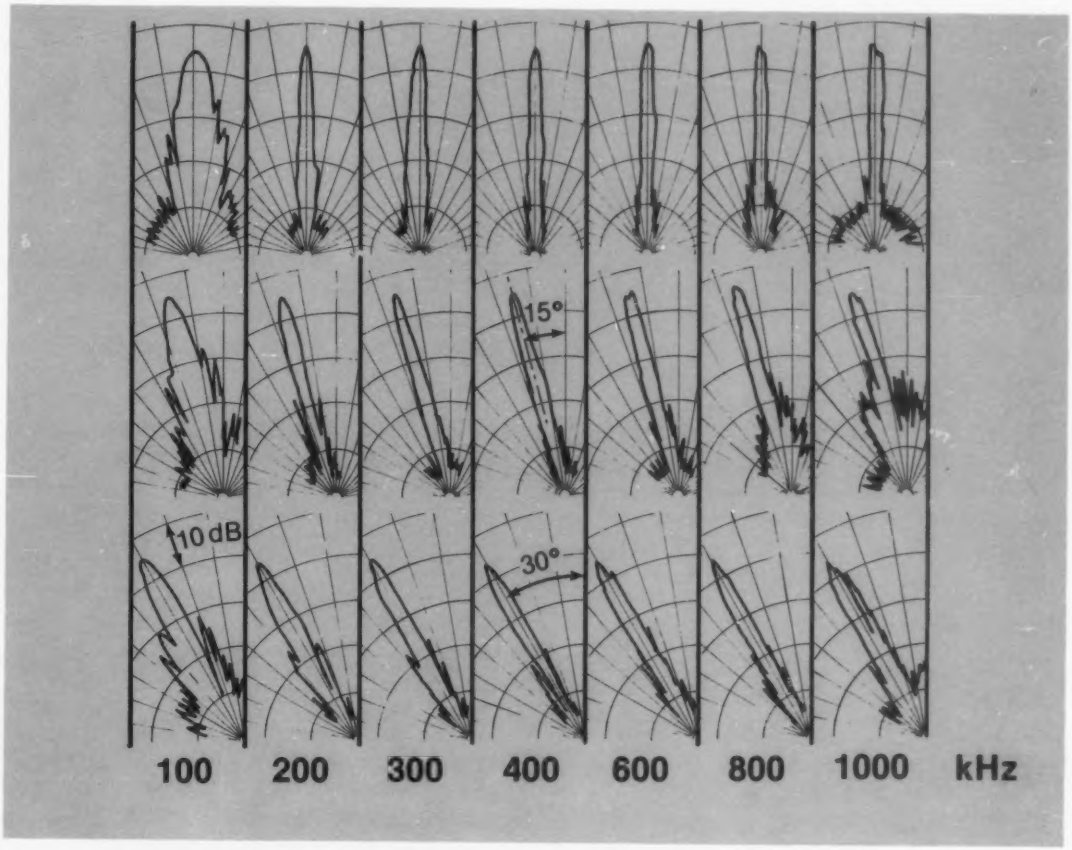


Figure 9. *Experimentally measured frequency-independent acoustic lens-acoustic filter plate receiving beam patterns taken at various frequencies and scan angles showing approximately constant beamwidths of about 2° at frequencies above 100 kilohertz.*

The array antenna concept studied by Rogers and Van Buren consists of an appropriately shaded spherical-cap transducer, or array of separate transducer elements closely approximating such a spherical-cap, as indicated in Figure 10(a). In this concept, if the angular extension of the spherical-cap transducer or array is $2a$ as illustrated, and the shading is such that the distribution of the normal

and to have an essentially constant beamwidth independent of the frequency whenever the latter exceeds a certain cutoff value. The transducer cap or array is mounted on a rigid acoustically-opaque sphere and is apparently intended to be used primarily as a transmitter although the same principles would clearly permit its use as a receiver.

In the related concept envisioned by Jarzynski and Trott the antenna is made up of an array of isophase, omnidirectional transducer elements mounted on a spherical surface much as before except that here—in principle although not in practical approximations—the array should completely fill out the sphere as in Figure 10(b). With shading again ex-

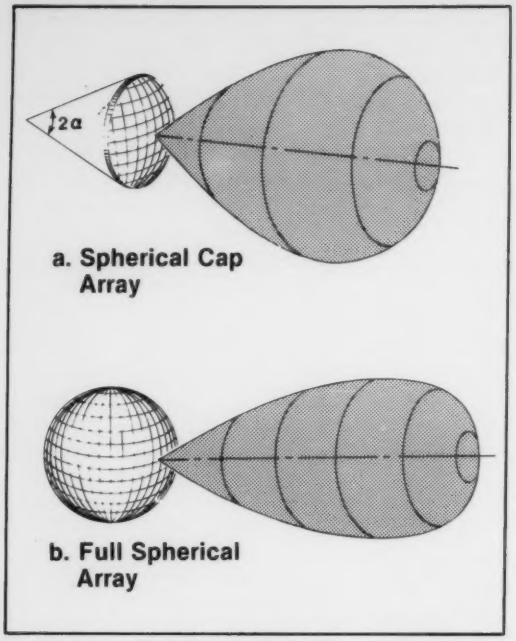


Figure 10. Spherical cap and full spherical array concepts for frequency-independent constant beamwidth beam formation; (a) spherical cap array of Rogers and Van Buren and (b) full spherical array of Jarzynski and Trott.

pressed in terms of Legendre functions the beam formed is again theoretically free from sidelobes and has essentially constant beamwidth independent of frequency above a certain cutoff value. A unique and essential difference between this concept and that of Rogers and Van Buren is that here the sphere on which the transducer elements are mounted is required to be acoustically transparent, a criterion perhaps not easy to realize with typical transducer hardware. As before, the concept may be applied in principle to either transmission or reception.

Although no successful experimental work with either version of the spherical array device has yet been reported, clearly the method, particularly in the form studied by Rogers and Van Buren, offers several interesting attributes including a potential—via suitable switching of the shading networks—for steerable frequency-independent, constant beamwidth beamforming for either transmission or reception.

A New Mosaic-Lens Array Antenna

Perhaps the most recent notion in the development of antennas of the type reviewed in this paper is the new "mosaic" lens, time delayed, antenna array concept of Sternberg and Anderson (12) reported

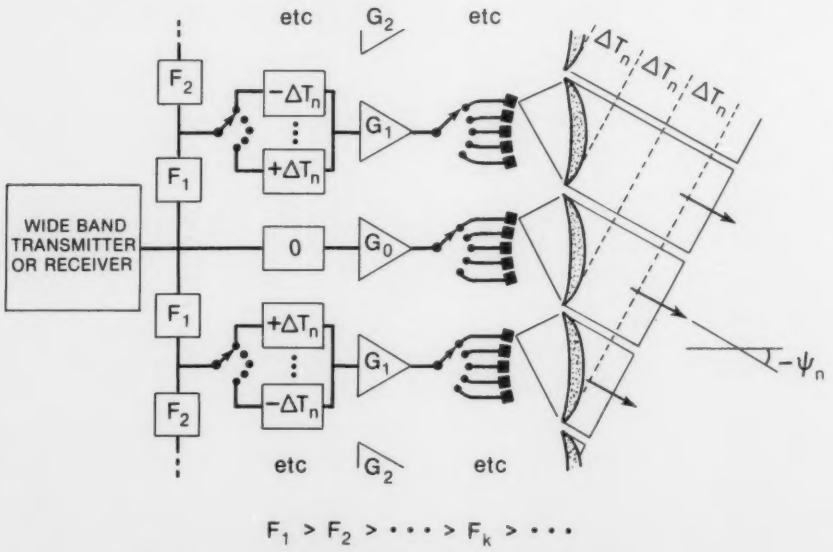


Figure 11. Mosaic-lens array antenna concept of Sternberg and Anderson for multiple or scannable approximately frequency-independent, approximately constant beamwidth beam formation suitable for large scale applications.

here in the open literature for the first time. In this approach to the problem, shown schematically in Figure 11, a linear or two-dimensional array of fully directive antenna elements—in fact, lens antennas each of which is accompanied with a suitable retina of transmitting and/or receiving elements—is controlled by one or more arrays of auxiliary switches, amplifiers, time-delay units and ideal low-pass filters as illustrated. Here, effectively heavy shading is provided by ideal low-pass filters, such that the highest frequency terms are effectively passed by the antenna array only through the central lens or a central group of lenses. The next highest frequency terms are passed in addition through those lenses which are immediately adjacent to the central lens or central group, and so forth. As a consequence of this shading, the effective aperture diameter D_E of the entire antenna array varies with frequency so as to maintain the ratio $R = \lambda/D_E$ approximately constant, as a result of which the beamwidth $B.W. \cong 65\lambda/D_E$ will also be maintained approximately constant. At the same time, with suitably prescribed positive or negative incremental time-delays, ΔT_n, applied to the signals passing through each lens outside the central lens as suggested in the illustration, beam steering independent of the frequency may be readily accomplished. The significant point here is simply that, because the antenna array elements are fully-directive lens antennas rather than the usual omni-directional point sources, the incremental time-delays required depend only on the center-to-center lens spacing, or lens diameter, L, for equal diameter lenses as in the illillustration, and on the beam steering angles, Ψ_n, as shown. In particular, the time-delays do not depend on the signal wavelength λ. Thus, for the arrangement illustrated, the required incremental time-delays are given by the formula

$$\Delta T_n = \frac{L}{c} \sin \Psi_n$$

where c is the sound velocity in the case of sonar. Phase shift in the ideal low-pass filters must be reduced to a negligible level and the transducing elements in each retina must be appropriately matched in primary pattern widths to each lens, over the band of

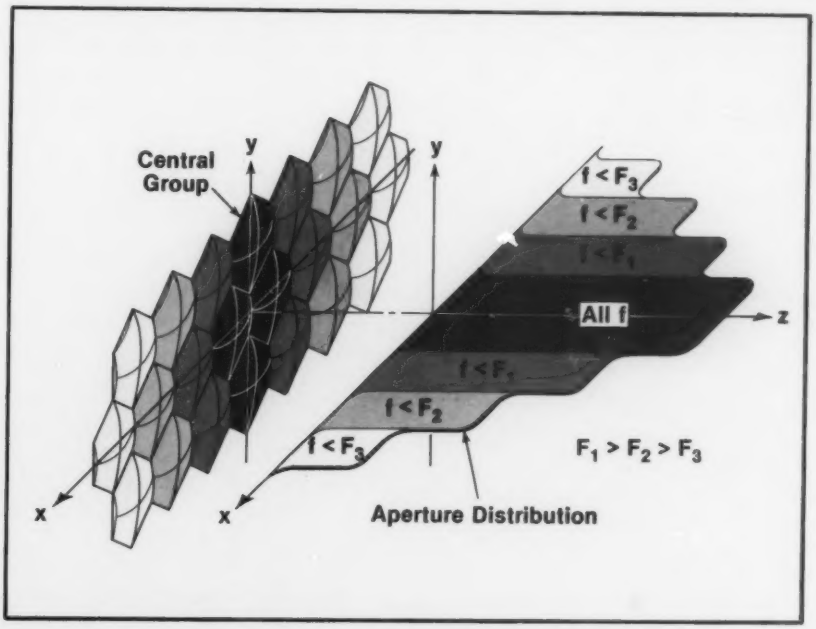


Figure 12. Mosaic-lens array antenna concept for approximately frequency-independent fan beam formation using hexagonally-cut lenses with a central group of four lenses showing the approximate aperture spectral distribution and shading function when the fan beam is pointed in the z-axis direction and fanned in the y, z-plane.

frequencies to be covered.

Moreover, with the arrangement shown in Figure 11, i.e., with a suitable controlling array of auxiliary switches, amplifiers, time-delay units and ideal low-pass filters, not only may a single, approximately frequency-independent, constant beamwidth steerable, transmitting or receiving beam be formed by the antenna, but, with additional "banks" of controlling arrays of auxiliary switches, amplifiers and so forth—which in the illustration may be thought of as "piled up" in layers all connected through the switching devices to the same array of lens antenna elements—many such approximately frequency- independent, constant beamwidth beams may be formed simultaneously in many different directions at once. In addition, if desired, the multiple beams thus formed may also be separately steered by the antenna.

Finally, many variations in the "mosiac-like" arrangement of the fully directive lens antenna elements may be used. One of the simplest of these is that shown in Figure 12, where hexagonally-cut lenses are mounted closely together so as to form an approximately frequency-independent, directional beam with approximately constant beamwidth in the plane of scan, or azimuth plane of the array antenna, as illustrated.

Some General Conclusions

It is clear that frequency-independent, constant beamwidth beamforming offers a number of significant advantages in sonar and probably also in radar applications and that on a single steerable beam basis it can be done in a number of different ways. Moreover it is clear that several of the methods available may be readily extended to provide for multiple beam formation, i.e., for the simultaneous formation of many such beams in many directions without the necessity for time sharing. While several of the methods appear better suited to transmission than to reception, or vice versa, others are equally applicable to each mode of operation or indeed to both with the same antenna.

Of the several methods discussed here the lens-filter plate approach for passive sonar reception appears distinctly different from, and somewhat simpler than the others, in that the beam formation is accomplished in an analog fashion without elaborate demands on control or signal processing devices. However, because of obvious problems connected with fabricating the lenses and filter plates if built more than a meter or two in diameter or a few meters across at most, it is quite clear that the lens-filter plate approach to the problem is practical only at frequencies at least above one or even 10 kilohertz. For example, to reproduce the patterns of Figure 3 in the spectral band from say 2.5 kilohertz to 25 kilohertz the lens and filter plate would each have to be made about two meters in diameter and the filter plate alone would weigh about 5,000 pounds. To reproduce the patterns of Figures 8 and 9 in the spectral band from say one kilohertz to 10 kilohertz the lenses and filter plates would each have to be some 15 to 20 meters in diameter and the filter plates would each use about 250 tons or more of stainless steel! Thus further research and develoment in this direction can be expected to remain largely restricted as at present to high frequency bands only, but for these it has much to recommend it.

On the other hand it seems equally clear that all of the array approaches to the problem, including the mosaic-lens antenna array method, will require a great deal of control or processing electronics whether used for transmission or reception, with the scale of those requirements rapidly increasing for high resolution applications. Of the array methods described, none has been extensively investigated experimentally although each appears to offer its own special advantages and disadvantages. Thus, for active transmission, the spherical cap approach of Rogers and Van Buren should be capable of handling both reasonably high power and scanning by use of suitable switching matrices, but since only a cap of the sphere is used to form any one beam the full aperature of the overall antenna is always larger than that on which the beam is based at any given time. Contrariwise, although the mosaic-lens antenna array scheme of the present writers eliminates this deficiency and uses the full antenna aperture at all times, it would very likely be somewhat limited in power density transmitting capability. Research and develoment in connection with these concepts will therefore very likely be governed by the special requirements and limitations imposed by the intended area of application and system cost factors in the larger scale cases where no other solutions are viable.

In summary, very few completely final conclusions can be drawn at this stage about the pros and cons of the various methods thus far proposed for doing frequency-independent beamforming. It would be fair to say, however, that for high-resolution reception at relatively high frequencies the lens-filter plate approach has many advantages to offer in sonar technology. But if narrow beams and high directivity are not required, or if frequency-independent, constant beamwidth beamforming is to be done at low frequencies for either transmission or reception—say

perhaps for oceanographic research or geophysical sub-bottom exploration—then one of the several array methods described would most probably be the better choice and will drive the research and development process.

Finally, in closing we would like to suggest that although both the spherical array work of Rogers and Van Buren and the lens-filter plate approach of the present authors are currently active Navy programs likely to continue for some time, it is our belief that additional scrutiny is in order of all of the methods of frequency-independent beamforming proposed to date, and especially, that more extensive experimental work on the most promising array methods needs to be carried out. ■

References

1. Dyson, J.D., "The Equiangular Spiral Antenna," *I.R.E. Transactions on Antennas and Propagation, 7,* 1959, 181-187.

2. Dyson, J.D., "The Unidirectional Equiangular Spiral Antenna," *I.R.E. Transactions on Antennas and Propagation, 7,* 1959, 229-334.

3. Tucker, D.G., and Gazey, B.K. *Applied Underwater Acoustics,* Pergamon Press, Oxford and New York, 1966, 20-21.

4. Rumsey, V.H., *Frequency Independent Antennas,* Academic Press, New York and London, 1966, 1-145.

5. Morris, J.C. and Hands, E., "Constant Beamwidth Arrays for Wide Frequency Bands," *Acoustica, 11,* 1961, 341-347.

6. Morris, J.C., "Broad Band Constant Beamwidth Transducers," *Journal of Sound and Vibration, 1,* 1964, 28-40.

7. Hixson, E.L. and Au, K.T., *Broadband Constant Beamwidth Acoustical Arrays,* Acoustic Research Laboratory, University of Texas at Austin, T.M. 19, May 1970, 1-37.

8. Sternberg, R.L., Anderson, W.A. and Stevens, G.T., "Log-Periodic Acoustic Lens-Acoustic Filter Plate Study," *Journal of the Acoustical Society of America, 63,* May 1978, 1617-1621.

9. Rogers, P.H. and Van Buren, A.L., "New Approach to a Constant Beamwidth Transducer," *Journal of the Acoustical Society of America, 64,* July 1978, 38-43.

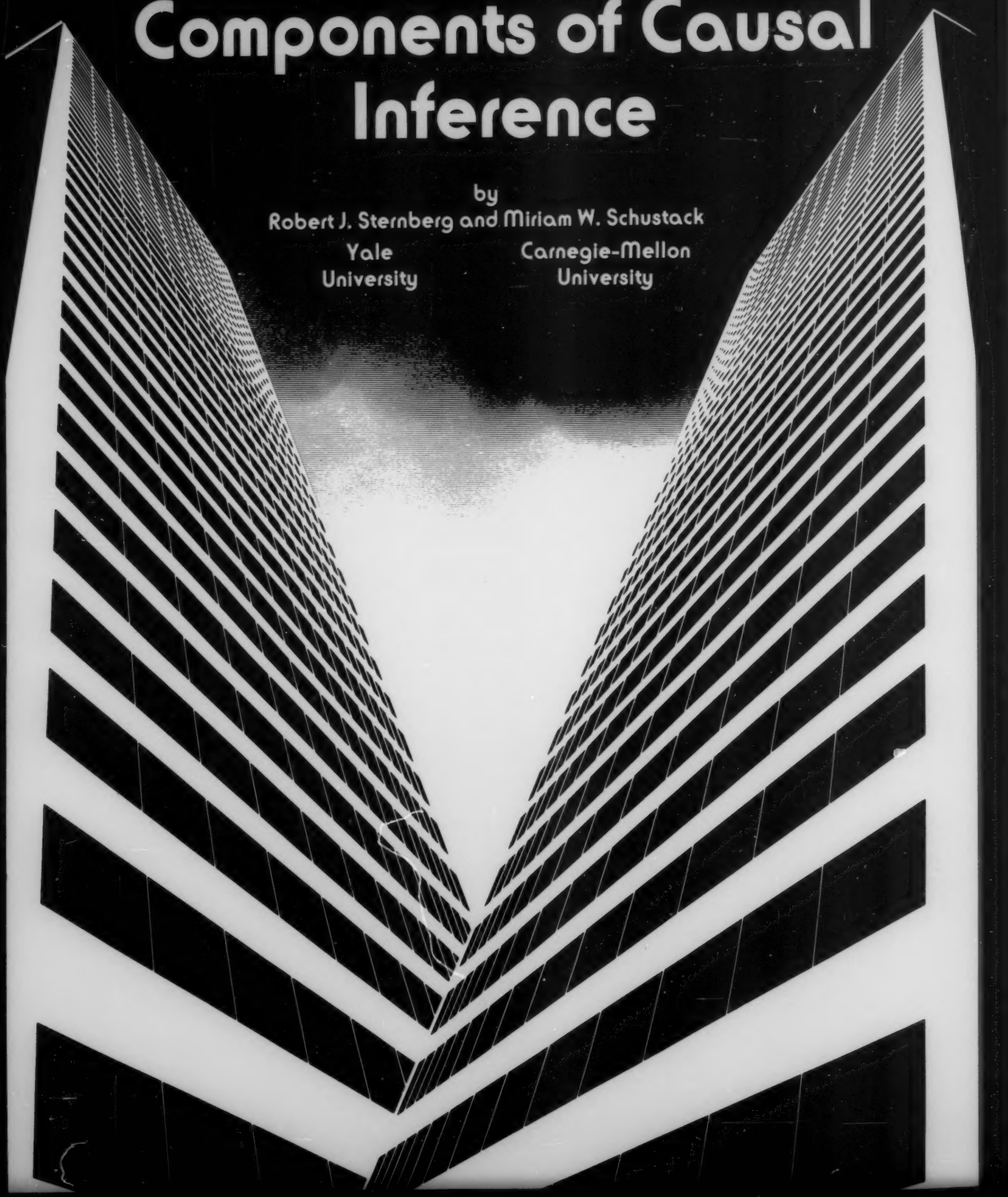
10. Jarzynski, J. and Trott, W.J., "Array Shading for a Broadband Constant Directivity Transducer," *Journal of the Acoustical Society of America, 64,* November 1978, 1266-1269.

11. Tucker, D.G., "Arrays With Constant Beamwidth Over a Wide Frequency Range," *Nature, 180,* 1957, 496-497.

12. Sternberg, R.L. and Anderson, W.A., *Frequency Independent Beamforming,* Office of Naval Research-Naval Underwater Systems Center Note, May 1979, 1-5.

Components of Causal Inference

by
Robert J. Sternberg and Miriam W. Schustack
Yale University
Carnegie-Mellon University



The Componential Analysis of Human Intelligence

The research reported in the accompanying article represents one study in a program of research at Yale on cognitive information processing in human intelligence. This research, which is supported by the Office of Naval Research, utilizes a set of techniques called "componential analysis," the object of which is to isolate the basic information-processing components that constitute an important part of intelligent behavior, and to relate these components to each other and to performance on intelligence tests and on real-world tasks.

Componential analyses of human intelligence seek to answer several questions about intelligent behavior: First, what are the information-processing components people use in intelligent performance? Second, how do people combine these components into working strategies for task performance? Third, upon what kinds of internal representations do these processing components and strategies act? Fourth, how quickly and accurately are the various components executed? And finally, how do people differ in their processes, strategies, representations, speeds, and accuracies of component execution?

Answers to these questions enable us to advance toward two major goals. The first is the formulation of a general information-processing theory of intelligence. This theory, described in Sternberg (1, 2, 3), attempts to show how intelligent behavior can be understood in terms of a relatively small number of information-processing components that enter into a large variety of test-like and real-world tasks. The second goal is the formulation and implementation of programs for training intelligent performance. These programs, described in Sternberg (4), have so far enabled us to improve people's skills in high-level reasoning tasks, such as reasoning by analogy, and in high-level reading tasks, such as the acquisition of meanings of unfamiliar words from natural contexts.

References

1. *Sternberg, R. J. The nature of mental abilities. American Psychologist,* **1979,** *34, 214-230.*

2. *Sternberg, R. J. Stalking the I.Q. Quark. Psychology Today, 1979, 13, 42-54.*

3. *Sternberg, R. J. Sketch of a componential subtheory of human intelligence. Behavioral and Brain Sciences, in press.*

4. *Sternberg, R. J., Ketron, J. L., and Powell, J. S. Componential approaches to the training of intelligent performance. Intelligence, in press.*

Civilian and military personnel alike frequently need to infer the causual antecedent or antecedents of significant real-world events. For example, the response of the United States government to the recent Soviet invasion of Afghanistan was predicated upon the motives U.S. government officials inferred to underlie the Soviet attack. The response would almost certainly have been different if government policy-makers had believed Soviet claims that the Soviets were merely responding to a request from the Afghan government for help in resolving domestic turmoil.

Series of casual inferences tend to form casual chains that can have major consequences for the course of world events. The Soviet counter-response to the American response to the invasion of Afghanistan, for example, was predicated upon the motives the Soviets inferred to underlie the American response, and so on. More generally, American (or Soviet) interventions in other parts of the world, whether military or otherwise, have been in large part a function of policy-makers' inferences regarding the causes of international upheavals.

The importance of casual inference, at the in-

Dr. Robert J. Sternberg** is Associate Professor of Psychology at Yale University. He received his Ph.D. in 1975 from Stanford University, where he received the Sidney Siegel Memorial Award for his dissertation on the topic of human intelligence. Dr. Sternberg has authored and edited several books and numerous articles dealing with human intelligence, and is associate editor of the journal, **Intelligence.

***Miriam W. Schustack** is a graduate student at Carnegie-Mellon University, where she is completing her doctoral work in cognitive psychology. She received her bachelor's degree from Princeton University, where she majored in psychology. Ms. Schustack's research interests are in causal inference, language comprehension, learning by analogy, and the development of linguistic abilities in children.*

dividual as well as at the governmental level, is such that it is crucial to understand how people evaluate evidence in making causal inferences. In particular, it is important to know in what ways the evaluation of evidence departs from a strictly rational model of evaluation, so that decision-makers can be made aware of these departures, and perhaps eventually be trained to make more rational inferences, or at least to take their own irrationality into account.

What Is A "Cause"?

What exactly is meant by the "cause" of a certain event? There seems to be no clear consensus in everyday usage of the term, and even philosophers have disagreed as to its precise meaning. There are three basic schools of thought which exist regarding the meaning of the term (1). These schools of thought are not mutually exclusive, in that the word *cause* is almost certainly used in multiple ways to convey somewhat different meanings. The first school of thought argues that a cause is a *necessary condition.* In other words, if a particular event does not occur, the consequent will never occur either. For example, the presence of electrical-impulse conduction in the brain is a necessary condition of human life as we know it. The second school of thought argues that a cause is a *sufficient condition.* In other words, if a particular event occurs, the consequent is sure to occur as well. For example, the absence of electrical-impulse conduction in the brain is sufficient for "brain-death." The third school of thought, which might be viewed as a "pragmatic" one, views a cause as a *condition that makes a critical difference.* It defines a cause as that "incident or action which, in the presence of those conditions that usually prevail, made the difference between the occurrence or nonoccurrence of the event" (1, pp 399-401). If, for example, a detective wished to find out why a person died, he or she would probably describe as "causal" the fact that a bullet was found in a homicide victim's brain, not the fact that the electrical impulses of the victim's brain ceased to be conducted.

Optimal Causal Inference

The philosophical analysis of causal inference dates back at least to John Stuart Mill (2), who proposed a set of "canons" for inferring causality. Mill's goal, however, was to propose optimal strategies for causal inference, rather than to model the strategies people actually use in making causal judgments. Mill proposed five strategies, of which we shall briefly consider two. In the Method of Agreement, if two or more instances of the phenomenon under investigation have only one circumstance in common, that circumstance is the cause of the given

phenomenon. For example, if some outcome, O, results from events A, B, C, and D; and if the same outcome results from events A, E, F, and G; then one infers that A is the cause of the outcome. In the Method of Difference, if an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common except one, then that circumstance is the cause, or an indispensable part of the cause, of the phenomenon. For example, if some outcome, O, results from events A, B, C, and D; and if the same outcome fails to materialize as a result of events B, C, and D; then one concludes that A is the cause, or an indispensable part of the cause of O[1].

Real-Life Causal Inference

People don't always follow an "ideal" or normative strategy in causal inference. Recent research on causal inference has been concerned with how people actually make causal judgments, rather than with how they should make them. Much of this research has been social psychological in nature, with the goal of understanding how people respond to social-psychological variables. A common finding, for example, is that people are more inclined to ascribe their own failings as due to external factors (for example, bad luck), but to ascribe the failings of others to internal factors (for example, the others' incompetence). One of the leading social-psychological investigators, Kelly (4), has suggested that attributions of causality can be viewed in a systematic framework in which effects on causal inference of different kinds of internal and external factors can be isolated and identified.[2]

A Cognitive Approach To Studying Causal Inference

Our own research attempts to specify the kinds of evidence individuals consider in evaluating the likelihood that a given outcome event is the result of a particular hypothesized causal event. Individuals must make these judgments with incomplete information about complex problems that vary simultaneously on many dimensions. In making judgments, individuals must decide what kinds of evidence to consider, how to weigh each kind, how to combine various types, and how to translate their conclusions into a probability that the target hypothesis was responsible for the target outcome.

The problems used in our research resembled in many respects problems encountered in real-world settings that require causal inferences.[3] We used three basic kinds of problems that differed in the content domain in which the causal inference is to be made. The three kinds of problems are illustrated in Table I by: (1) an epidemiological domain, in which the individual had to judge the likelihood that a particular hazard was responsible for a given epidemic; (2) a securities domain, in which the individual had to judge the likelihood that a particular circumstance was responsible for a precipitous decline in the value of a company stock; and (3) an abstract domain, in which the individual had to judge the likelihood that a particular circumstance (labeled only by a letter) was responsible for some other particular circumstance (also labeled only by a letter).

In each problem of each kind, individuals were presented with the hypothesis that a particular event was responsible for some outcome. They were asked to use a given body of evidence to estimate the probability (placed on a 0-to-100 scale to eliminate decimal points) with which that event, by itself, would produce that outcome. Individuals were explicitly warned that they were being given incomplete information and that interactions between possibly causal events were possible; these warnings were aimed at evoking the same kind of mental set as would real-life causal inference. In the problems with concrete (epidemiological or securities) content, each city or company represented a "situation"; in the problems with abstract content, each line represented a situation. In all problems, the result (to the right of the arrow in the abstract-content condition) was the occurrence or nonoccurrence of the outcome event. The hypothesized cause (the "target") to be evaluated was always specified.

Within any situation in any kind of problem, each possible cause was in one of three states: observed to be *present* (e.g., "The office staff of the company organized and joined a union", or the occurrence of event "R" in the abstract example); observed to be *absent* (e.g., "The office staff of the company did not organize or join a union", or the

[1] *Further information regarding Mill's Canons can be obtained from Copi (1) and from Skyrms (3).*

[2] *Further information regarding social-psycholigical approaches to causal inference can be found by consulting Carroll and Siegler (5), Chapman and Chapman (6), Fischhoff (7), Gollob, Rossman, and Abelson (8), Jones, Kanouse, Kelley, Nisbett, Valins, and Weinter (9), Kelley (10), Nisbett and Ross (11), and Tversky and Khneman (12).*

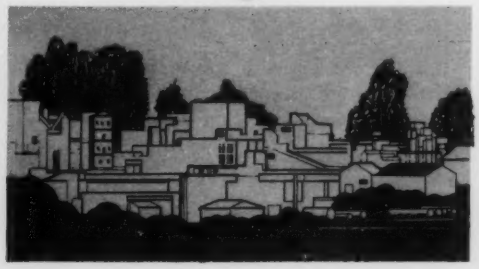
[3] *More detailed reports of this research can be found in Schustack and Sternberg (13).*

Table I
An Example of a Causal Inference Problem in each of Three Content Domains

Epidemiological Domain

An epidemiologist noted that, for Marshall-Isaacs disease:

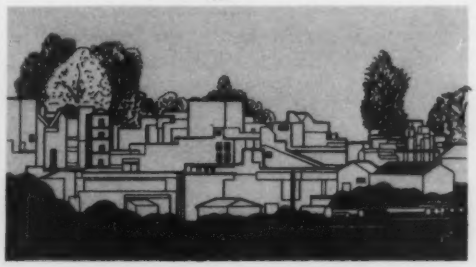
City 1.



- Annual health inspection of food-service workers was stopped.
- A new type of hair dye was introduced to the area.
- No inadequate sterilization practices were used in a local cannery.
- A new type of pesticide was tried by local vegetable farmers.
- There was a water-main break.

AN EPIDEMIC OF THE DISEASE WAS REPORTED.

City 2.



- Inadequate sterilization practices were used in a local cannery.
- Annual health inspection of food-service workers was not stopped.
- Lead-based paint was sold without warning labels.
- There was no water-main break.
- A new type of pesticide was tried by local vegetable farmers.

NO EPIDEMIC OF THE DISEASE WAS REPORTED.

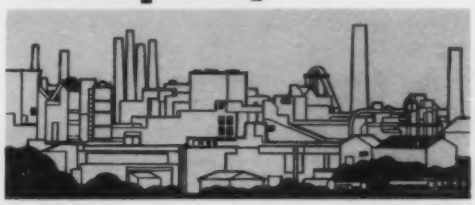
What is the probability that, in some other city, there would be an epidemic of Marshall-Isaacs disease if a new type of pesticide were to be tried by local vegetable farmers?

Table I Continued

Securities Domain

A market analyst noted that, among pharmaceutical manufacturers:

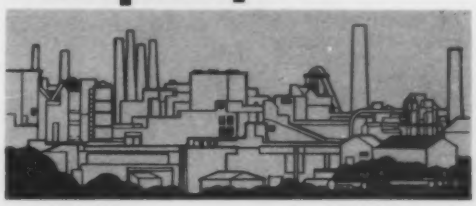
Company 1.



- The office staff of the company organized and joined a union.
- The company's major product was under suspicion as a carcinogen.

THERE WAS A DRASTIC DROP IN THE VALUE OF THE COMPANY'S STOCK.

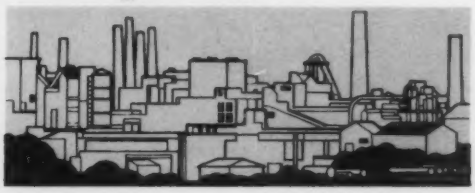
Company 2.



- The office staff of the company did not organize or join a union.
- The company's major product was under suspicion as a carcinogen.

THERE WAS A DRASTIC DROP IN THE VALUE OF THE COMPANY'S STOCK.

Company 3.



- Illegal campaign contributions were traced to the company's managers.
- The company's major product was not under suspicion as a carcinogen.

THERE WAS NOT A DRASTIC DROP IN THE VALUE OF THE COMPANY'S STOCK.

What is the probability that, for some other pharmaceutical manufacturer, stock values would drop drastically if the company's major product were under suspicion as a carcinogen?

Abstract Domain

R T H Q → B

Q –R L T → –B

–T –H R Q → –B

L Q –R T → B

–Q –T U H → –B

R → B

absence of event "R" in the abstract example); or *not observed* (e.g., nothing is stated about unionization or nothing is stated about "R"). Over the set of problems in the three content domains there were from two to five situations (cities, companies, or lines of problems) described in a single problem, and also from two to five possibly causal events observed per situation independently of the number of situations. Within a single problem, each situation had the same number of observed possibly-causal events.

Sixty-two subjects supplied probability ratings (on the 0-100 scale) for each of the two concrete content domains. Half of those subjects rated first epidemics and then securities; the other half rated securities first and then epidemics. Forty subjects supplied probability ratings for the abstract content domain only. An additional 21 subjects supplied "base rate" ratings for the concrete content domains that were intended to tell us the likelihood that the different outcomes would be obtained without explicit information. The base-rate problems were the same as the concrete-content problems, except that no evidence was provided. Subjects were asked to judge the probability of the conclusion solely on the basis of what they knew from their own past experience. Subjects in this condition saw questions such as "What is the probability that, in some city, there would be an epidemic of Phipp's Disorder if diseased cattle were slaughtered and sold?" and "What is the probability that, for some coffee company, stock values would drop drastically if there was political upheaval at the company's foreign location?" All subjects were members of the Yale University community.

Theory Of Components Used In Decision-Making In Causal Inference

How do people go about this task? How do they decide how likely it is that a possible cause leads to a particular consequence? Such decisions would seem to be quite difficult to make, especially in light of the incomplete and sometimes seemingly incompatible information available both in real life and in our causal inference problems. We propose that people use four basic kinds of information regardless of item content as summarized in Table II:

1. Confirmation by joint presence of possibly causal event and outcome. For example, in Line 1 of the abstract problem in Table 1, "R" is observed to be present, and the outcome of interest, "B", is observed to be present as well. This relation between the possibly causal event and the outcome event is evidence *for sufficiency* of the possibly causal event for the outcome, i.e., if the possibly causal event occurs, so does the outcome.

2. Confirmation by joint absence of possibly causal event and outcome. For example, in Line 2 of the abstract problem in Table 1, "R" is observed to be absent, as indicated by "-R" before the arrow, and "B" is also observed to be absent, as indicated by "-B" after the arrow. This relation between the possibly causal event and the outcome event is evidence *for necessity* of the possibly causal event for the outcome, i.e., if the possibly causal event does not occur, neither does the outcome.

3. Disconfirmation by presence of possibly causal event but absence of outcome. For example, in Line 3 of the abstract problem in Table 1, "R" is observed to be present, but the outcome of interest, "B", is absent, as indicated by "-B". This relation is evidence *against sufficiency* of the possibly causal event for the outcome, i.e., the occurrence of the possibly causal event does not always lead to the occurrence of the outcome.

4. Disconfirmation by absence of possibly causal event but presence of outcome. For example, in Line 4 of the abstract problem in Table 1, "R" is absent, as indicated by "-R" before the arrow, but the outcome of interest, "B", is present. This relation is evidence *against necessity* of the possibly causal event for the outcome, i.e., the occurrence of the outcome is not always preceded by the occurrence of the possibly causal event.

It was also possible in our experiments for there to be no information regarding the possibly causal event, despite there being information (about presence or absence) regarding the outcome of interest. For example, in Line 5 of the abstract problem in Table 1, there is no observation regarding "R", but the outcome of interest is observed to be absent ("-B"). Such lack of information about a particular situation was theorized to have no significant impact upon people's judgments.

A fifth kind of information, base rate of occurrence of the outcome of interest, applies only to problems with concrete content. In these problems, it refers to the probability of the outcomes occurring in the absence of any new information regarding the probability of occurrence in a particular new situation. We theorized that there might be some effect of base rate, but that if there were, it would be small. A large literature provides examples of situations in which base rates are possibly relevant, but are nevertheless ignored (e.g., Lyon & Slovic, 14; Nisbett, Borgida,

If event is	and outcome is	this is evidence that possibly causal event is
Present	Present	Sufficient
Absent	Absent	Necessary
Present	Absent	Not Sufficieint
Absent	Present	Not Necessary
No Information	Present or Absent	No Significant Effect

Table II
Four Basic Kinds of Information

Crandall, & Reed, 15). Following Nisbett and Borgida (16), we suggest that when people have a great dcal of other information they tend to ignore base rates.

In our examples, we have shown how the components of decision-making described above could be applied to the "target" event, i.e., to the possibly causal event about which a judgment is to be made. We theorized that these components would be applied to "competitive" possibly causal events as well, i.e., to other possible causes that might lead to the outcome of interest, instead of or in addition to the target event. Preliminary data analyses suggested that people took into account only the two "strongest" alternative hypotheses, i.e., the two possibly causal alternatives having the most evidence to support them.

Combination of Components

We hypothesized that the four kinds of information (components of causal inference) enumerated above are combined in a simple, additive way; in other words we assumed that each time evidence of a given kind appeared in a particular problem, a subject would count it as one additional piece of evidence of that kind for or against the likelihood that the target event led to the outcome of interest. We made predictions, about the probabilities that would be assigned by experimental subjects, via linear equations in which the dependent variable was the observed probability that the target event in isolation led to the outcome of interest, and the independent variables were the numbers of occurrences of each kind of evidence. A parameter was estimated for each independent variable that reflected the weight assigned to evidence of that particular type. For example, in the epidemiological problem there is one example of confirmation by joint presence of antecedent and outcome: a new type of pesticide was tried by local vegetable farmers in City 1, and an epidemic of Marshall-Isaacs disease was reported in City 1. The independent variable used for prediction of the effect of confirmation by joint presence was therefore "1", and the parameter estimated for that independent variable corresponded to the weight assigned to this piece of evidence in making the final probability judgment. An additive constant in the linear equation represented the "base probability" constant across all problems with a given kind of content. For the alternative hypotheses, only the two strongest alternatives were considered, and a single independent variable was formed by subtracting, from the average number of occurrences of evidence tending to support the two alternatives, the average number of occurrences of evidence tending to refute these two alternatives. Thus, the standard form of the equations used in parameter estimation was

Probability of target event leading to outcome of interest =

(Number of confirmation by joint presence

of target event and outcome) x (weight assigned to this number)
+ (Number of confirmations by joint absence of target event and outcome) x (Weight assigned to this number)
+ (Number of disconfirmations by presence of target event but absence of outcome) x (Weight assigned to this number)
+ (Number of disconfirmations by absence of target event but presence of outcome) x (Weight assigned to this number)
+ (Average number of confirmations minus average number of disconfirmations of two strongest alternative antecedents) x (Weight assigned to this number)
+ (Base probability constant across all items of one content).

In our particular experiment, we used 60 problem types with one equation corresponding to each of the problem types. The 60 problems varied in the amounts of evidence of each kind tending to support or refute the target event. Each of the 60 problem structures appeared once for each content; thus, a given problem appeared in three different forms, with content distinguishing each form from the other two forms.

Goals of Research

Our research on causal inference was intended to discover how people evaluate evidence in determining how likely it is that a given event leads to a particular outcome. To this end, we wished to

1. test whether the four components of causal inference described above, when applied both to the target event and to competitive alternative events, provide a sufficient or nearly sufficient account of how people decide how likely it is that a given event, in isolation, results in a certain outcome. We also planned to compare alternative accounts of causal inference, ones that differed from our proposed one in the components used, and in the ways in which these components are combined;

2. examine effects of base rates in problems with concrete content;

3. determine how much people weighted each of the components of causal inference in arriving at their judgments of likelihood, and compare the weightings of components across content types;

4. compare the usage of components, combination rules, base rates, and weightings of components across individuals, finding major sources of individual differences (if there were any to be found);

5. consider the implications of our findings for the training of decision-makers, exploring, in particular, whether any kinds of evidence that should be considered are not considered, and whether any kinds of evidence that should not be considered are considered.

Results

Basic Statistics

Table III shows basic statistics for the response probabilities obtained for each of the three kinds of stimulus content. The difference between the mean probabilities for concrete content (averaged over epidemics and securities) and abstract content was not statistically significant, $t(100) = .62$, p .10. This is consistent with the notion that subjects in both the abstract and concrete content groups were using similar or identical decision rules in making their judgments. A matched-pairs t-test was performed on the means of subjects' responses for epidemics and securities. It was found that the mean for epidemics was significantly lower than that for securities, $t(61) = 2.87$, p .01. Given that the mean estimate of probabilities for the base rates was also significantly lower for the epidemics (33.96) than for the securities (44.65), the difference may be attributable to the base-rate differences between the two kinds of content, rather than to differences in the decision rule applied.

Having considered these basic statistics, we now turn to a consideration of the questions expressed by the basic goals of the experiment.

Fit of Theory to Data

Fit of the proposed theory (as quantified above) to the experimental data was assessed in several ways. Each of these ways provides somewhat different information about the relation between the predictions of the theory and the observations obtained from the subjects.

First, a squared correlation (R^2) was computed for each content type between the predictions of the theory and the observations obtained from the subjects. This value represents the proportion of observed variation in responses across problem types that the proposed theory is able to explain. A value of 0 would indicate no fit of the theory to the data; a value of 1 would indicate perfect fit. Values of R^2 were .90, .88, and .90 for epidemic, securities, and abstract content respectively. All of these values of R^2 obviously are quite respectable, indicating that the theory did very well in predicting subjects' response patterns. It is important to note, however, that the variance in the dependent variable not explained by

	Epidemics	Content Securities	Abstract
Mean Response	35.16	37.25	34.99
Standard Deviation (across item types)	20.92	18.96	18.90
Minimum Item Mean	10.05	10.44	10.45
Maximum Item Mean	76.35	75.40	74.44

Table III
Basic Statistics

the theory was also statistically significant in each case, indicating that although the fits of theory to data were quite good, they were not perfect: there was systematic variation in the data that was not accounted for.

Second, the root-mean-square deviation (RMSD) of observed from predicted values was calculated for each data set. This value measures the extent to which predicted values differ from observed values in the original metric of the responses. Since responses were on a scale from 0 to 100, an RMSD of 0 would indicate that the predicted means did not differ at all from the observed values; a value of 100, possible only in the limiting case, would indicate that the predicted values differed maximally from the observed values. Note that for RMSD, unlike for R^2, lower values indicate better fit. The values of RMSD were 6.72, 6.39, and 6.04 for the epidemic, securities, and abstract content respectively. These values indicate that the predicted values were quite close to the observed ones—differing on the (root-mean-square) average by only about 6 points.

Third, the predictions of the proposed theory were compared to the predictions of psychologically plausible alternative theories. The idea behind each of these alternative theories was that people do not simply count instances, but rather consider instances in relation to some known, larger quantity. These alternative theories were of two basic types. One type proportionalized the number of instances of each kind of evidence over various sums, namely, (a) the total number of situations (cities, companies, or lines) in each given problem; (b) the number of situations in each given problem in which information about the target event (either presence or absence) was provided; (c) the number of situations with the same outcome (e.g., number of situations with joint presence of target event and outcome of interest divided by total number of situations of any kind in which the same outcome, presence, occurs); (d) a combined version of (b) and (c); (e) the number of possibly causal events for the outcome of interest. None of these alternative theories fit the data as well as did our original formulation, however. The ranges in values of R^2 for the five alternatives were from .72 to .85 for the epidemics (with a median of .78), from .75 to .87 for the securities (with a median of .80), and from .72 to .86 for the abstract problems (with a median of .76). There was thus no reason to favor any of these more complicated decision rules over the original one.

The second type of alternative model tested was nonlinear. Nonlinear models assuming interactions among various components were fit to the data, but none resulted in nontrivial improvement in fit over the simpler additive model without interactions.

On the basis of these results, we concluded that our original four-components model, supplemented by a base probability constant, fit the data very well, and better than the alternatives we tested.

Effects of Base Rates

Base rate (the likelihood that different outcomes would be obtained without explicit information) was added as an additional independent variable to the prediction equation presented earlier. The result was

a nonsignificant improvement in R^2 for the securities problems, but a significant improvement ($p < .001$) for the epidemics. Although this latter improvement was statistically significant, it was not very large, accounting for less than .02 increment in R^2. Clearly, the incremental impact of base rates was small. One possible explanation for this near-neglect of base rates by subjects is that the use of two different concrete content domains led them to believe that they should treat even the concrete problems abstractly and disregard base rates, despite our instructions that encouraged the use of prior information about causality. This explanation probably does not account for our findings, however. Assuming that any possible contamination due to multiple content domains would affect the second-presented set of items exclusively (or at least primarily), we would then expect base rates to be more important for the first-presented set of items for each subject. This was not the case, however.

Weightings of Components

Table IV presents parameter estimates for each of the six estimated parameters (four components for the target antecedent, the combined component for the competitor alternative antecedents, and the base probability constant) for each of the content conditions. Both the signs and magnitudes of the parameter estimates are strikingly consistent across the three content conditions. These similarities in values are further evidence of the modest use of base rate, since base rates for individual items were not matched across problems, and in the abstract condition base rates were nonexistent. The signs of the parameter estimates provide further plausibility for the proposed theory, in that variables that should have increased the probability of the target event leading to the outcome of interest did increase this probability (positive sign), and variables that should have decreased the probability of the target event leading to the outcome of interest did decrease this probability (negative sign).

All parameter values for variables relating to the *sufficiency* of the target event for the outcome have higher values than all parameter values for variables relating to the *necessity* of the target for the outcome. In the case of joint presence and joint absence, evidence in favor of sufficiency versus necessity is confounded with numbers of negations (0 vs 2 negations for the two respective types of evidence). But in the case of mixed presence and absence, evidence against sufficiency and necessity is not confounded with numbers of negations (1 and 1 for the two

respective types of evidence). It appears, therefore, that subjects viewed sufficiency information as more important than necessity information in causal determinations of the kind required in our experiment.

The relatively low value assigned to the composite variable measuring strength of the two strongest alternative events came as something of a surprise. Apparently, the very fact that an event was selected as a target (for no known particular reason) made that target event much more salient in subjects' decisions than alternative events that happened not to be selected.

Finally, it is of interest to notice the relatively high value assigned to the base constant parameter. This value is constant across problems of a given content type, and corresponds to the probability value that would be assigned in the absence of any information. The high values of the constants show again the large amount of "credit" an event received merely by virtue of its having been selected as the target. In the absence of any information at all, the target would theoretically be assigned a probability of about 35 (on the 0 to 100 scale).

Individual Differences

Our attempts to find consistent differences among subjects were unsuccessful. In fact, the performance of subjects on the causal inference task was notable for its uniformity. Reliabilities of the responses were computed by examining coefficient α across all possible splits of subjects (rather than the usual way of examining it across all possible splits of items). A high value of α would indicate high uniformity of data patterns across subjects. Reliabilities of .98, .97, and .99 were obtained for epidemics, securities, and abstract problems respectively, indicating very highly consistent patterns of responses across individual subjects. Reliabilities computed across earlier trials (first half of items) and later trials (second half of trials) were also in the high .90's.

We fit the proposed theory to the data of each individual subject. The mean value of R^2 for the individual-subject data was .64 (with a standard deviation across subjects of .15). Since the reliability of individual-subject data averaged .83, we were accounting for .64/.83, or .77 of the reliability in the individual-subject data. We attempted to use the individual model fits to find subjects who were likely to be using decision rules other than the proposed one that were not merely idiosyncratic to a single subject.

	Content		
	Epidemics	Securities	Abstract
Joint Presence of Target and Outcome (In favor of sufficiency)	11.30	9.22	9.26
Joint Absence of Target and Outcome (In favor of necessity)	3.79	2.52	2.87
Presence of Target but Absence of Outcome (Against sufficiency)	-8.42	-10.05	-8.17
Absence of Target but Presence of Outcome (Against necessity)	-6.44	-5.75	-8.07
Strength of Alternatives (Two Strongest)	-3.20	-3.30	-3.15
Base Probability (Constant)	29.96	36.20	33.60

Table IV
Parameter Estimates

Our strategy for doing this was to identify the six subjects for whom the proposed theory accounted for the lowest proportion of *systematic* variation in response-choice data. These were subjects who were going about the problems sytematically, but who were presumably following a system somewhat different from the one we expected. Unfortunately, the mean correlation between response patterns for these subjects was only .34, which suggested that these subjects were *not* following any consistent decision rule; correlations between pairs of subjects' data were uniformly low. This correlation of .34 can be contrasted with one of .64 between the data sets of those subjects for whom our theory accounted for the highest proportions of systematic variance in their data. To conclude, the subjects poorly fit by our theory appear to have been using idiosyncratic or possibly inconsistent decision rules. There was no evidence for their use of some other shared decision rule.

We also correlated individual parameter estimates for the proposed theory with subjects' scores on three psychometric tests of reasoning ability. There were no significant correlations between parameters and tests, suggesting that subjects' weightings of various evidence types were not systematically related to their reasoning abilities.

Implications for Training

The results presented above indicate that people are able to deal effectively with complex information about causal relationships, and to assign highly reliable estimates of probabilities of causal connections between variables on the basis of evidence that is incomplete and that may be interactive in unknown ways. The high reliability of the data is particularly striking in view of people's reported difficulty and frustration in performing the causal inference task we presented: subjects found the task to be extremely challenging, and generally reported they were not aware of having used any systematic strategy for solving the problems.

Although people showed high correspondence with *each other* in their strategies for causal inference, they did not show high correspondence with what one might conceive of as "ideal decision maker." Thus there may be utility in the training of individuals to improve their skills of causal inference. Such training would focus upon several aspects of decision-making.

First, people grossly undervalue evidence about alternative hypotheses. The target hypothesis received an enormous advantage merely by virtue of its having been selected (in these problems, arbitrarily) as the target. It appears that once people adopt a certain causal hypothesis they become strikingly resistant to strong evidence supporting alternative causal hypotheses. This tendency is evident at all times in scientific theorizing, where new theories are almost always at a competitive disadvantage simply by virtue of their being new. The tendency could be and probably has been most unfortunate in policy-making decisions. For example, one can see without too much effort the tendency of governmental decision-makers to interpret actions of Communist governments in terms of a current and consensually validated world view of what these governments ultimately have in mind. During times of Cold War, actions of Communist governments tend to be perceived as hostile and self-serving in devious ways; during times of detente, the same actions are more likely to be perceived as reasonable and ultimately non-threatening. What is disturbing about this fact is that the current world view is simply too likely to appear to be confirmed, regardless of the set of events that transpires. Many of the recent mispredictions of events abroad, for example, such as the overthrow of the Shah of Iran, seem to have resulted from resistance on the part of policy-makers to evidence that tended to contradict their world views of the time. People need to be trained to take greater heed of plausible alternative causal antecedents for outcomes of interest.

Second, people greatly overvalue positive confirmations (joint presence of target and outcome) relate to other types of evidence. The parameter corresponding to joint presence generally had the highest weight for the group data (of the four components considered), and it also had the highest weight for 90% of the individual subjects. Our finding of a bias toward positive confirmation of previous hypotheses is consistent with earlier reseach in concept attainment and reasoning (e.g., Wason 17, 18; Wason & Johnson-Laird, 19). People need to be trained to pay greater attention to the other three kinds of components. These other kinds of evidence can be helpful in establishing causal links, but they are not given a fair chance when people make their causal inferences.

Third, people underutilize base-rate information (see also Tversky & Kahneman, 12, 20). In many instances, particularly in those instances in which new information is unreliable, base rate information can provide an important supplement to what is newly learned. Doctors, for example, learn to use base rates to assign low prior probabilities to the interpretation

of symptoms in terms of exotic diseases if the symptoms fit a less exotic disease equally well. Obviously, though, over-reliance on base rates can lead to blindness to the importance of new information that does indeed suggest, for example, the presence of an exotic disease.

Fourth, people essentially ignore information about co-occurrence of alternative causal events with the target one. Suppose, for example, that in one problem there is evidence supportive at a certain level of a given event, A, having caused a given outcome, 0. Now suppose that in another problem, the evidence for event A is equally favorable, but there is another event, B, that occurs whenever A does. The co-occurrence between the two events should reduce people's estimates that either event, in isolation, leads to the outcome of interest. The reason is that the evidence might actually be supporting only A, only B, or both A and B, either individually or collectively. One simply cannot tell for sure, since A and B are completely confounded in the information available. In our experiment, people were utterly oblivious to information about co-occurrence of potentially causal events. People must be trained to spot confoundings in evidence, and to take them into account when they make their causal inferences.

Finally, people probably need to be sensitized to the implications of missing information and of potential interactions among potentially causal variables. A simple, linear model taking into account only available evidence probably can provide a reasonable first-order estimate of the probability of a causal connection. Such an estimate is only an initial one, however: missing information, and nonlinearities such as those caused by interactions, will need to be taken into account in order to obtain more sophisticated estimates. If information is particularly scarce, for example, people need to be trained to delay decision-making until sufficient evidence is available.

In Conclusion

Our experiment provided us with some initial evidence regarding the components people use in making causal inferences, the decision rules they use in combining these components, the weights they assign to the components, and the ways in which the causal inferences people make differ from those they might make if they were more sophisticated decision-makers. We have also performed two other experiments (See Schustack & Sternberg, 13, Note 1), which supported the conclusions we have drawn for (a) a set of abstract problems that was not isomorphic to those used in the present experiment and for (b) a set of abstract problems containing only confirming evidence (of the kinds represented by the first two components described earlier).

Our current plans include an investigation of the following topics: causal inference in situations where people are asked to judge probabilities of causal connections for multiple possibly causal events, rather than for single possibly causal events in isolation; how causal inference develops with age; and the effects of social-psychological variables (e.g., inferences about the self versus inferences about others; inferences dealing with environmental variables versus inferences dealing with psychological variables). We believe that this line of research is potentially relevant to high-level civilian and military decision-making, and we hope that in time we or others will be able to apply findings of the kind presented here to the training of decision-makers of all kinds. ■

References

1. Copi, I. M., *Introduction to Logic* (fifth edition), Macmillan, New York, 1978.

2. Mill, J. S., *A System of Logic, Ratiocinative and Inductive,* J. W. Parker, London, 1843.

3. Skyrms, B., *Choice and Chance: An Introduction to Inductive Logic* (second edition), Dickenson, Encino, California, 1975.

4. Kelley, H. H., "Attribution Theory in Social Psychology," in D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15), University of Nebraska Press, Lincoln, Nebraska, 1967.

5. Carroll, J. S., & Siegler, R. S. "Strategies for The Use of Base-rate Information," *Organizational Behavior and Human Performance,* 1977, *19,* 392-402.

6. Chapman, L. J., & Chapman, J. P., "Illusory Correlation in Observational Report," *Journal of Verbal Learning and Verbal Behavior,* 1967, *6,* 151-155.

7. Fischhoff, B., "Attribution Theory and Judgment under Uncertainty," in J. H. Havey, W. J. Ickes, & R. F. Kidd (Eds.), *New*

Directions in Attribution Research, Erlbaum, Hillsdale, N.J., 1976.

8. Gollob, H. F., Rossman, B. B., & Abelson, R. P., "Social Inference as a Function of The Number of Instances and Consistency of Information Presented," *Journal of Personality and Social Psychology,* 1973, *27,* 19-33.

9. Jones, E. E., Kanouse, D. E., Kelley, H. H., Nisbett, R. E., Valins, S., & Weiner, E. (Eds.), *Attribution: Perceiving the Causes of Behavior,* General Learning Press, Morristown, N.J., 1972.

10. Kelley, H. H., *Attribution in Social Interaction,* General Learning Press, Morristown, N.J., 1972.

11. Nisbett, R. E., & Ross, L., *Human Inferences: Strategies and Shortcomings in Social Judgment,* Prentice-Hall, Englewood Cliffs, N.J., 1980.

12. Tversky, A., & Kahneman, D., "Causal schemata in judgments under uncertainty," M. Fishbein (Ed.), *Progress in Social Psychology,* Erlbaum, Hillsdale, N.J., 1977.

13. Schustack, M. W., & Sternberg, R. J., "Evaluation of Evidence in Causal Inference," *Journal of Experimental Psychology: General,* in press.

14. Lyon, D., & Slovic, P., "Dominance of Accuracy Information and Neglect of Base Rates in Probability Estimation.' *Acta Psychologica,* 1976, *40,* 287-298.

15. Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H., "Popular Induction: Information is Not Necessarily Informative," J. S. Carroll & J. W. Payne (Eds.), *Cognition and Social Behavior,* Erlbaum, Hillsdale, N.J., 1976.

16. Nisbett, R. E., & Borgida, E., "Attribution and the Psychology of Prediction," *Journal of Personality and Social Psychology,* 1975m *32,* 932-943.

17. Wason, P. C., "On the Failure to Eliminate Hypotheses in a Conceptual Task," *Quarterly Journal of Experimental Psychology,* 1960, *12,* 129-140.

18. Wason, P. C., "On The Failure to Eliminate Hypotheses . . .—A Second Look," in P. C. Wason & P. N. Johnson-Laird (Eds.), *Thinking and Reasoning,* Penguin, Harmondsworth, England, 1968.

19. Wason, P. C., & Johnson-Laird, P. N., *Psychology of Reasoning: Structure and Content,* D. T. Batsford, London, 1972.

20. Tversky, A., & Kahneman, D., "Judgment under Uncertainty: Heuristics and Biases," *Science,* 1974, *185,* 1124-1131.

research notes

Modeling of Sound Scattering By Seamounts

In several regions of the world's oceans, seamounts represent the major features of bottom bathymetry, and in these cases where their height is a significant fraction of the ocean water depth, they can significantly affect the propagation of underwater sound. To determine the degree of shadowing and its effect on sound propagation, Canadian investigators at the Defense Research Establishment Pacific conducted a series of propagation runs at a frequency of 230 Hz at the Dickens seamount in the Gulf of Alaska. Their results, presented at the Fall 1978 meeting of the Acoustical Society of America, confirmed simple ray theory arguments that the most significant effect of propagation occurs when the source is shallow and most of the rays are deeply refracted. In order to quantify the result of the experiment, ONR principal investigator Dr. Herman Medwin and his graduate student Bob Spaulding, Jr., at the Naval Postgraduate School have developed a physical model. To account for diffraction, the investigators constructed a detailed three-dimensional model of Dickens seamount from laminated layers of particle board and simluated the at-sea experiment in an anechoic chamber. The total propagation loss was calculated by adding laboratory model values of upslope forward scatter and crest diffraction losses to computer-predicted ray refraction losses up to and away from the seamount. The laboratory predictions were found to agree almost exactly with the field data for the fully blocked case thus showing that the diffracted energy over complex barriers can be predicted utilizing physical laboratory models. ■

(J. M. McKisic, ONR)

Scanning Acoustic Microscope to be Commercially Produced

Over the past several years, Professor C. F. Quate and collaborators at Stanford University have developed scanning acoustic microscopy to a sophisticated and useful technique. Various phases of Professor Quate's work have been supported at various times by the Hartford Foundation, DARPA, AFOSR, NIH, JSEP, NSF, and ONR.

Those years of research are now being applied to the production of a commercial acoustic microscope by E. Leitz-Wetzlar GmbH of West Germany. Leitz manufactures high-quality optical microscopes which are used extensively in the semiconductor industry. K. Heitmann, director of Leitz-Wetzlar's acoustic microscope project, expects to have a prototype completed within one year. During this year, there will be a continuous exchange of information between the Stanford team and the manufacturer in Germany. Negotiations are under way to have some of the components produced by U.S. electronics manufacturers.

Simply stated, a scanning acoustic microscope (SAM) is constructed in such a way that an object is mechanically scanned through the narrow waist of a focused beam of high frequency sound to obtain a "picture" of the acoustic reflectivity (or transmissivity of the object). At room temperatures, the resolution attained is about 0.52 microns (comparable to green light in a good optical microscope). However, it is important to realize that, aside from resolution, the picture obtained from differences in acoustic parameters is different

from the optical picture, and in many instances more informative. For example, subsurface structure may be revealed. ■

(Logan E. Hargrove, ONR)

Prostaglandins and Oxygen Toxicity

Elevated partial pressures of oxygen (pO_2) in breathing-gas mixtures for divers offer a great advantage in shortening decompression times by replacing the non-metabolizable inert gas by one that is quickly metabolizable. Unfortunately, high pO_2 is toxic to the lungs and the central nervous system so that the pO_2 used must be carefully controlled to avoid serious consequences. The amelioration of this restriction would provide a ready means for making decompression phases more efficient.

Prostaglandins (PG) are a family of substances found in many tissues of the body, particularly in the lungs when they are detoxified (i.e. broken down chemically and removed). One of the ONR Physiology Program investigators, Dr. Christopher Schatte, discovered that aspirin (a potent inhibitor of prostaglandin formation) increases the susceptibility of animals to oxygen toxicity. Suspecting that PG metabolism might be intimately involved in oxygen toxicity, he investigated the synthesis, circulating levels and breakdown of PG in a hyperoxic environment. He found that hyperoxia did not affect either the synthesis or the circulating blood levels of PG. The breakdown of the PG in the lungs (catabolism) was impaired as a result of a decrease in activity of the rate-limiting enzyme complex (PG dehydrogenase/reductase) in PG catabolism. If some of the PG compounds are proven to be causally related to pulmonary oxygen toxicity, susceptibility of the lungs to oxygen poisoning could possibly be altered by manipulating dietary fat content since fat is known to have an effect on PG synthesis. ■

(Leonard M. Libber, ONR)

Miniature Cryogenic Refrigerators

Small, energy-efficient closed cycle cryogenic refrigeration systems ("cryocoolers") for use with superconductive electronic sensors are being developed under a NAVMAT-funded ONR/NRL program. An application of great potential interest to the Navy for such cryocoolers is their deployment with superconductive magnetometers and gradiometers in low frequency surveillance and communication systems. In response to a requested design study for a small cryocooler with defined detailed specifications, three out of eight companies were successful bidders. They were SHE Corporation, North American Phillips Corporation and CTI Corporation. It is anticipated that one or more of the design studies will be used in a Phase II endeavor to build prototype units to couple with superconducting magnetometer systems.

For possible Magnetic Anomaly Detection (MAD) applications, the magnetic signature of the cryocooler must be sufficiently small so as not to degrade the performance of the superconductive detector. NRL plans to establish a test and evaluation facility to measure the magnetic signatures of several cryocooler systems to be developed under the ONR/NRL Program. ■

(Edgar A. Edelsack, ONR)

Electromagnetic Signatures of Arctic Sea Ice

Under contract to ONR, Dr. R. K. Moore of the University of Kansas Remote Sensing Laboratory utilized an L-band and Ku-X-band scatterometer to obtain electromagnetic signatures of various types of sea ice under winter conditions, which further demonstrated the ability of microwave radar to discriminate ice types. During these experiments, the effects of varying conditions of snow cover on the electromagnetic signatures of ice under winter conditions were also determined. Thick first-year, multi-year, and pressure-ridged sea ice and fresh-water lake ice were investigated at 1.5

GHz and from 8-18 GHz, at angles of incidence from 10° to 70° (with respect to nadir), and at both parallel and cross antenna polarizations. The L-band (equivalent frequency to SEASAT SAR) radar was found to have the ability to discriminate among fresh water ice, the thick categories of sea ice, and pressure-ridged sea ice. This work confirms the practical use of L-band radar for topographical mapping.

Discrimination of ice types was found to be better at the Ku-X-band frequencies. The greatest separation between the two ice types of perhaps greatest concern, thick first-year and multi-year sea ice, was found to occur at 9 GHz. Significant levels of separation existed between angles of incidence from 15° to 70°.

On effects of snow cover on radar signatures of thick first-year sea ice and freshwater lake ice, it was found that at L-band frequencies, snow cover of 0 to 8 cm had very little effect. At the 8-18 GHz frequencies, the contribution of snow cover to the overall signatures became increasingly more pronounced with increasing frequency. Snow roughness was also found to have a significant effect in certain cases. At the lower X-band frequencies, rough snow cover resulted in dramatically higher signatures at the larger angles of incidence.

This work contributes to our ability to interpret remote-sensing imagery of sea ice, and to derive specific parameters from sea ice which aid in predicting sea ice conditions. ■

(Charles A. Luther, ONR)

Navy Summer Faculty Research Program

Twenty-six science and engineering faculty members from colleges and universities throughout the United States have spent the summer at Navy laboratories in cooperative research with Navy scientists under a Navy Summer Faculty Research program. Sponsored and monitored by the Office of Naval Research, the program was administered by the American Society for Engineering Education (ASEE). The 26 participants in the 1980 program, now in its second year, were selected from among 400 faculty members who applied. Six of the participants worked at the Naval Reserach Laboratory in Washington, D.C.; the remaining participants were assigned among other major Navy R&D Centers, including Naval Weapons Center, Naval Ocean Systems Center, Naval Underwater Systems Center, David W. Taylor Naval Ship Research and Development Center, Naval Surface Weapons Center and the Naval Medical Research Institute.

The program affords selected college faculty members the opportunity to work with Navy scientists on current Navy research projects and to participate fully in the research and development activities of the laboratory in which they indicate an interest.

The objectives of this program are to involve university faculty members in the research programs of the Navy and to develop the basis for continuing research of interest to the Navy at the institution of the faculty member. It is also expected that this cooperative research will lead to the establishment of continuing relations among faculty members and their professional peers in the Navy and the enhancement of research interest and capabilities of the faculty members. ■

(Nicholas Perrone, ONR)

Tougher Ceramic Materials

Scientists at the Naval Research Laboratory have developed a new technique to produce ceramic materials that are tougher, more resistant to thermal shock, less prone to fracture, and more resistant to weather corrosion than conventional ceramic composites.

Such composites are expected to be used in the future by the Navy, and other DoD agencies, for structural materials in the fabrication of radomes and antenna windows, missile and rocket components, and other related structures which encounter high temperatures during use and are not exposed to long periods of oxidation at extreme temperatures.

The improved ceramic material is fabricated through the use of powder metallurgy techniques. The new technique calls for a mixture comprised of boron nitride particles, which are uniformly dispersed throughout an alumina matrix (Al_2O_3) with conventional wet-

powder processing techniques. The mixture is then hot-pressed into a ceramic composite exhibiting superior properties.

A patent is now pending on the invention. ■

(Roy W. Rice, NRL)

An Ultrasonic Non-Specular Reflection and Transmission Effects for Layered Media

Doctors T. D. K. Ngoc and W. G. Mayer, Physics Department, Georgetown University, have recently developed a numerical integration method under an ONR contract to calculate the intensity distribution in reflected and transmitted beams for liquid-solid-liquid layered media. The formulation describes all non-specular reflection and transmisssion phenomena for all angles of incidence, including critical angles. They found that the existence of non-specular phenomena depends not only on the angle of incidence and the product frequency times solid thickness but also on the product frequency times beam width. Calculations have been made for reflection and transmission and for incident angles corresponding to critical angles, plate-mode angles, and between-mode angles.

To evaluate the complex integrals it is essential that the imaginary part be non-zero. This condition is satisfied for all angles of incidence if the appropriate absorption coefficients for all the media are included. Previous methods are approximations which are, for example, valid only very near a critical angle. ■

(Logan E. Hargrove, ONR)

Solar Cycle Breakthrough

Dr. Robert Howard of the Mt. Wilson Observatory has made an important discovery that will revolutionize thinking on the solar dynamo and the nature of the solar cycle. He has discovered the existence of alternating zones of fast and slow rotation that originate at the solar poles and drift to the equator in 22 years (equal to the period of the Hale solar cycle). The pattern is symmetric about the equator, with amplitudes of about 3 meters per second. Solar activity zones are roughly centered on the upper boundary of the fast zones. The zone that will correspond to the next solar cycle (peaking in about 1991) has already originated at the pole during 1975.

The symmetry of the pattern suggests that we are observing the manifestation of a deep-seated phenomenon with a coherence scale the size of the sun. The observed pattern is associated with the subsurface fields that are responsible for the solar cycle.

This discovery is the first evidence of the association of mass motions with large scale characteristics of the solar cycle. It provides clear evidence that the solar cycle is not a random process generated by surface drifts of fields. Instead it is driven from beneath the surface by resonant large-scale mass motions. Professor Dicke has recently suggested that a perfectly periodic process deep in the interior of the sun drives the solar activity cycle, and that the apparent variability between cycles is caused by variations in the diffusion time from the interior to the surface. Dr. Howard's discovery is the first experimental evidence that suggests a deep interior source for the solar cycle.

It should be pointed out that this important fundamental discovery would not have been possible without improvements in resolution capability supported by ONR and, more importantly, without the 14 years of continuous support that ONR provided. ■

(Henry Mullaney, ONR)

Clarification of Sputtering Mechanisms

The process of sputtering, in which high velocity ions impinge on solid surfaces causing ejection of the substrate material, is used in various technologies to prepare thin films. Optical coatings and semiconductor films for electron devices are only two important areas of application. Another use of sputtering is in materials characterization, wherein the ejected

particles are analyzed to deduce the atomic nature of thin film materials as a function of depth from the surface. Despite these important applications, there is very little specific evidence which allows for a microscopic description of the mechanisms involved in the sputtering process. As a result it is not well controlled, nor are the important parameters understood.

Some recent work by Professor I. Tsong at Pennsylvania State University, funded by ONR, has revealed new facts about sputtering. Professor Tsong has carried out careful optical measurements on the decay of excited states of emitted atoms and ions. These measurements indicate that particles emitted from a variety of glassy materials have energies of about 100 to 1000 eV, which precludes any thermal excitation processes (a commonly used assumption). Professor Tsong further deduces that effects related to the electronic energy band structure of the solid do not appreciably influence the excitation process. This is counter to another proposed theory. It is apparent that the excitation is occurring in the first or second collision which occurs in the evolution of the sputtered atom. ■

(Larry R. Cooper, ONR)

ONR/NBS Cryocooler

The split-cycle Stirling cycle cryocooler being developed by Dr. James Zimmerman of National Bureau of Standards, Boulder, Colorado under contract with the Office of Naval Research, was the subject of a recent article in *Microwave Systems News.* In the Zimmerman-designed machine the moving parts near the cold head are made of epoxy, fiberglass and nylon to reduce magnetic contamination. Designed as a slow speed device of one cycle per second, its magnetic signature will be small enough to allow magnetometers to be operated with it. Using helium as a cooling medium, it has achieved a temperature of 6.5 K with milliwatts cooling power at that temperature. This capacity is adequate for several magnetometer configurations. At present the NBS refrigerators are in the laboratory prototype stage, however, serious consideration is being given to commercial development and production. ■

(Edgar A. Edelsack, ONR)

Fermi Level Pinning at Semiconductor/Liquid Junctions

Drs. Allen J. Bard, University of Texas and Mark S. Wrighton, Massachusetts Institute of Technology proposed the concept of Fermi level pinning at semiconductor/liquid junctions to explain the behavior of certain such junctions under illumination (*J. American Chemical Society, 102,* 3671 (1980)). This concept, resulting from research in their respective laboratories and supported, in part, by ONR under the program coordinated by the Navy Semiconductor Liquid Interfaces Committee, explains the observations that semiconductors such as n- or p-GaAs, and p-Si exhibit output photovoltages which are independent of the solution potential, even for electrolyte solutions containing redox couples having very different electrochemical potentials. Fermi level pinning has previously been used to explain the behavior of certain semiconductor-metal interfaces. The Fermi level pinning is proposed to result from surface states. Quantitative considerations indicate that surface state densities as low as $\sim 10^{12}$ cm^{-2} are sufficient to produce the pinning. As a result, the band bending is determined by the layer of surface states, not the electrolyte solution. Two significant consequences of Fermi level pinning may occur. First, considerable flexibility in the application of photoelectrochemical cells based on semiconductors exhibiting the pinning is lost by the fixing of the available photovoltage by the surface state layer-determined band bending. Second, and of a positive note, chemical processes previously thought impossible for energetic reasons may now be possible through manipulation of intermediate reactions to exploit the Fermi level pinning. For example, Bard has demonstrated photoinjection of electrons from p-GaAs into liquid ammonia although earlier simple theories predicted such a process would not occur. In addition, Wrighton has shown that species such as N,N′ -dimethyl-4,4′ -bipyridinium ion is photoreduced at p-Si in spite of the fact that the electrochemical potential for the reduction is more negative than the conduction band energy for p-Si. Finally, chemical modification of the semiconductor surfaces and, hence, the surface states offers a means of altering the photoelectrochemical properties. ■

(Jerry J. Smith, ONR)

Continued on page 72

A New Propulsor Concept for Ships

by
Herman E. Sheets
University of Rhode Island

A new concept for a ship propulsor has been developed at the University of Rhode Island under the sponsorship of the Maritime Administration. Data indicate that the concept is well-defined and offers a potential improvement in ship design, propulsion, and control. This program at Rhode Island, which began in 1974, was initially supported by the Office of Naval Research until the basic research phase was completed, and the project was ready for testing, evaluation and application.

The new propulsor uses a hydraulic transmission, located outside the ship's hull, driving the propeller. Thus, the reduction gear and its foundations are eliminated. The transmission uses seawater as the motive fluid, and the fluid discharge from the turbine contributes to the thrust of the propulsor. The propulsor consists of a prime mover (inboard) coupled directly to an (out-board) axial flow pump. Pressurized water is produced by the pump to drive a multistage turbine that is located in the hub of the

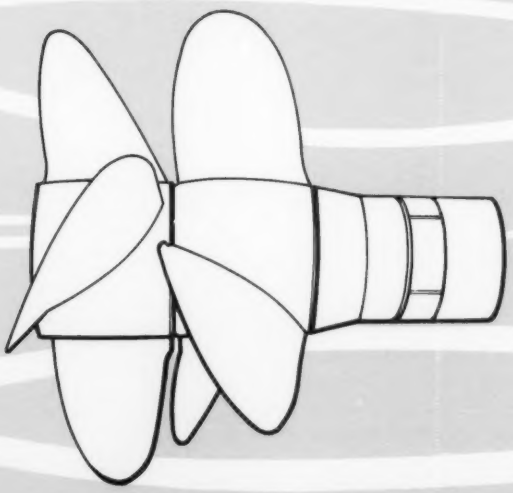
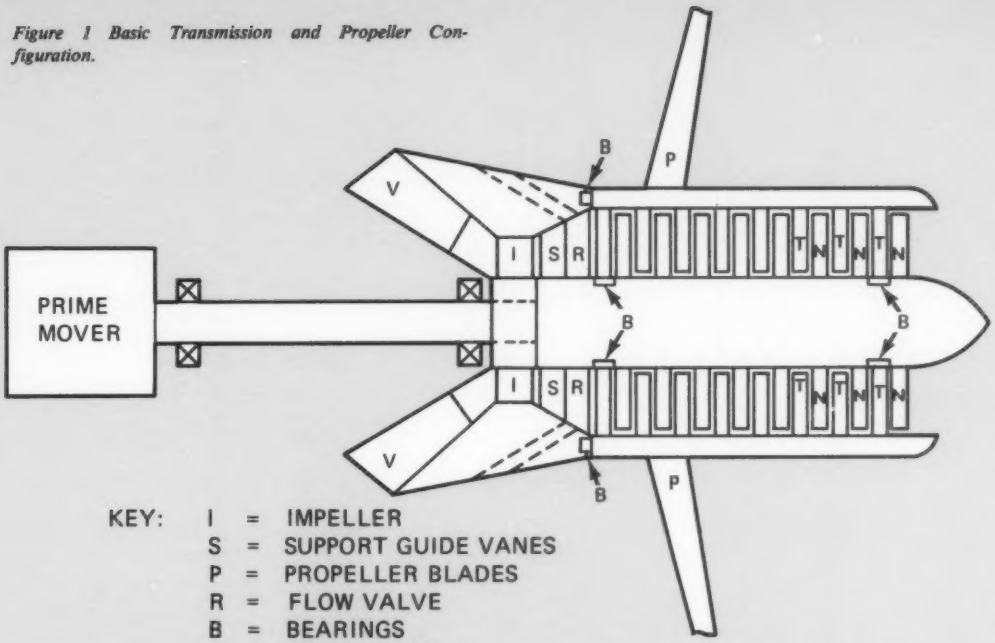


Figure 1 Basic Transmission and Propeller Configuration.



propeller and forms and integral part of the propeller. The entire weight of the transmission and propeller is of the same order of magnitude as the existing propeller for the same horsepower. There is a hydraulic thrust reversing mechanism incorporated in the transmission. The basic concept is shown in Figure 1.

Since the concept is entirely new, a proof-of-concept air model was built and successfully operated (with a 39:1 speed reduction ration) to demonstrate the structural principle of this system. A second air model was built shortly thereafter and successfully demonstrated the feasibility and simplicity of contrarotation propellers.

Dr. Sheets is Professor Emeritus of Ocean Engineering, University of Rhode Island and Chief Engineer for Analysis and Technology Inc., Stonington, Connecticut. He is a renowned pioneer in the field of fluid dynamics, especially in the area of advanced concepts for ship propulsion. He is a member of the National Academy of Engineering and the National Academy of Sciences.

With proof-of-concept established, a water model of 10 horsepower was built and operated to demonstrate the propulsor and to get experimental data on efficiency. The single-stage pump showed efficiencies between 85% and 90%. The five-stage turbine of the water model showed an efficiency of 91%. The entire transmission showed efficiencies of 71% with speed reductions on the order of 6:1 and up to 8:1.

On the basis of the test results and the experience gained during the engineering phase, a transmission system with 80% efficiency can be attained in the 10 horsepower size with a four-inch pump and a 5.7 inch turbine. The experimental data clearly indicated that a large transmission can be designed and built with efficiencies above 90%.

At the end of the program's third year, preliminary conclusions had been reached that this concept so far was more promising than initially anticipated. That feeling was reinforced by outside interest in the work and by presentations given and papers offered at scientific and development conferences where people from different backgrounds and disciplines were brought together. The work thus far did raise some questions about cost and efficiency that were examined.

An analysis was made to compare the new propulsor for ships with the standard gear transmission and the results indicate that the new propulsor equals, and with design improvements it can exceed, the efficiency of the standard gear transmission. Further improvement of the new propulsor over the standard gear transmission is the result of an increase in thrust due to the transmission turbine discharge flow. It became evident that the new propulsor can generate a higher thrust for the same propeller diameter, and it can produce a higher propeller efficiency for the same thrust. The new propulsor has weight savings on the order of 1.5 times the reduction gear weight and there are equivalent savings in space.

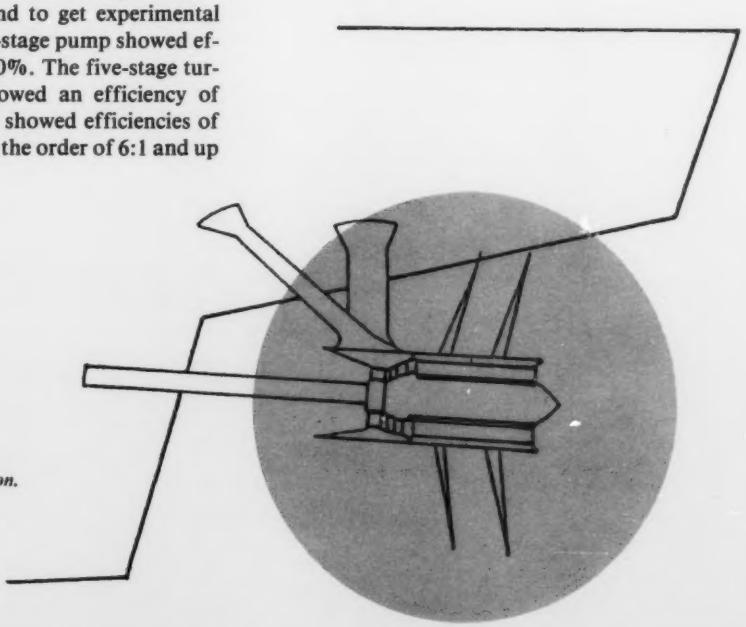


Figure 2

The Axial For Transmission used for Counter-Rotation Propeller Propulsion.

These savings can be used to improve ship's performance, increase pay load, or a combination of these factors.

A cost analysis has shown that this new concept will be less expensive than a gear transmission. This is easily understood due to the substantial reduction in weight of the propulsor compared to the gear transmission. The boundary layer control effects of the new propulsor on the ship's hull and the propeller have been assessed but not quantified, and they provide additional improvements in overall ship system efficiency.

The new concept has been studied, analyzed, and a research model has been built and tested. The studies have indicated that the optimum performance of the new concept should include the entire ship system, namely ship configuration and ship control in addition to ship propulsion. The transmission pump can be used for boundary layer control on the ship's hull and lead to entirely new and more efficient hull configurations. The transmission turbine discharge flow can be vectored in the desired direction, thus eliminating the ship's rudder. The concept is entirely new, and as with any new undertaking, additional improvements can be made with research on the reversing system, the transmission, as well as the integration of the propulsor into the ship.

Like any new concept its full realization is incremental, moving from one phase to another in a logical sequence. A tentative conclusion to be drawn from the work done thus far is that it is entirely possible to advance ship design and propulsion to a new level of performance through the implementation of a three-point program extending over three or four years consisting of 1) design and 2) build a scaled up model that will be 3) tested at sea. There is an abundance of research, analytical and experimental data to substantiate the validity of this recommendation as being the next logical step step to take. ■

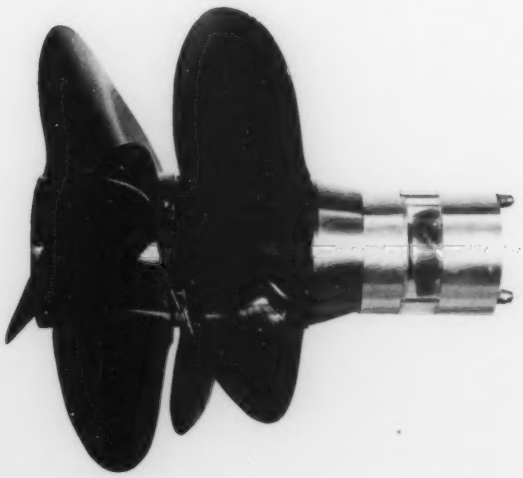


Figure 3 **Contra-rotating Air Model.**

Research Notes

continued from page 67

Transition of Adaptive Antenna Array Work

During the past year significant progress has been made in the adaptive antenna array work carried out at Ohio State University under its Joint Services Electronics Program. These arrays play an important role in interference-resistant communications. During the past year an improved least-mean-square (LMS) loop has been developed to eliminate time constant variation with signal power and hence improve the dynamic range of an adaptive antenna system. Also, the problem of grating lobes, which normally occur for widely-spaced elements, has been solved by properly selecting the element patterns in an adaptive array of widely-spaced elements. Polarization flexibility has been studied and found to enhance greatly the performance of an adaptive antenna system. Polarization control can be used to make a communications system much more resistant to intentional interference. Significant improvement also has been made in acquisition of a desired signal in the presence of interference by using a power optimization algorithm. This work has now moved beyond a basic study and further work is now supported by Naval Electronic Systems Command and Rome Air Development Center. ■

(John O. Dimmock, ONR)

Multilaser High-Speed Research Camera

A novel concept in high-speed photographic systems has been developed and demonstrated at the Naval Research Laboratory. The new system will permit quantitative measurements of ballistic penetration and other dynamic events vital to the Navy and other DoD agencies. Initial results have been obtained from ballistic penetration studies, photoelastic stress-wave analysis, and dynamic fracture experiments.

Six independent, Q-switched ruby lasers in the system provide extremely short (30 nanoseconds), yet intense, beams of light to record such events as ballistic penetration and stress wave propagation. A unique feature that enhances the system's versatility is the use of a microprocessor to control the timing and sequencing of actions in the laser cavity while the dynamic event is in progress.

The lasers and their high-voltage circuits are housed in an enclosed cabinet to provide for safe operation, and the laser light is transmitted to the rest of the optical system by fiber-optic cables to eliminate the problem of steering intense beams of light.

The system offers other advantages over conventional camera systems, such as a framing rate (0 to 2 million frames per second) variable over a wide-range. Other advantages include a light intensity greater than that from conventional sources, highly monochromatic light output (free of coherent optical noise) and system sharing made possible by the use of fiber-optic links to several experimental sites. ■

(Robert J. Sanford, NRL)

Revolutionary Tool for Biology and Medicine

Dr. Tom Manuccia and Dr. John Reintges of the Optical Sciences Division at the Naval Research Laboratory (NRL) are the first to show, using forefront laser technology, that it will be possible to identify and locate specific chemical bonds in living cells by coherent anti-strokes Raman emissions. A paper presented at the International Congress on Quantum Chemistry revealed their invention. Unlike classical Raman spectroscopy, their process uses two tuned lasers to drive or "pump" the specific type of chemical bond (C-H, C-D, D-O, C=O, etc.) under study to an almost continuous emission of visible light. The technique offers the first hope of nondestructive real-time-domain chemical localization, analysis, and quantitization in living and dividing cells. A continuous spectrum might be taken from a single cell undergoing division or a response to stimulus, or a specific compartment of a cell as small as the mitochrondria might be examined as a cell crawled across a glass slide. No other instrument offers these possibilities in biology today. NRL scientists are also excited by the potential industrial applications of the principles of this method. Potential patents on this invention would be wholly Navy owned. ■

(Paula T. Beall, ONR)

Monolithic Microwave Superhet Receiver

On the face of this coin are three chips of gallium arsenide semiconductor material containing the appropriate transistor, metallization, capacitors, and inductors for the critical ports of an X-bond superheterodyne receiver. They are, from left to right, (1) RF preomp, (2) mixer, and (3) IF buffer amplifier. Development in progress combines these chips into a single low cost, highly reliable monolithic circuit and adds the voltage-controlled oscillator function. The use of refractory (e.g., titontium: tungsten/gold) metallization and the complete elimination of flying lead wires significantly improves reliability. The totally monolithic construction reduces size and production cost by a factor of 100 below that of conventional approaches.

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
ARLINGTON, VA. 22217

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

POSTAGE AND FEES PAID
DEPARTMENT OF THE NAVY
DOD-316
CONTROLLED CIRCULATION RATE

U.S. MAIL